# A time series processing tool to extract climate-driven interannual vegetation dynamics using Ensemble Empirical Mode Decomposition (EEMD)

Pieter Hawinkel[a,b], Else Swinnen[a], Stef Lhermitte[b], Bruno Verbist[b], Jos Van Orshoven[b], Bart Muys[b]

[a]*Flemish Institute for Technological Research (VITO), Remote Sensing Unit, Boeretang 200, B-2400 MOL, Belgium*

[b]*KU Leuven – University of Leuven, Department of Earth and Environmental Sciences, Celestijnenlaan 200E, B-3001 Heverlee, Belgium*

**Abstract – Interannual changes of vegetation are crucial in understanding ecosystem dynamics under global change. However, there is no automated tool to extract these interannual changes from remote sensing time series. To fill this gap, the Ensemble Empirical Mode Decomposition (EEMD) framework was refined and implemented to decompose time series of Normalized Difference Vegetation Index (NDVI) and reconstruct their interannual components. The performance of EEMD-based interannual NDVI detection was assessed using simulated time series, and its sensitivity to model and data parameters was determined to provide a basis for remote sensing applications. The sensitivity analysis highlighted application limitations for time series with low interannual to annual amplitude ratios and high irregularity in timing of growing seasons, as these factors have the strongest effects on the overall performance. However, within these limitations, the detected interannual components correspond well to simulated input components with respect to timing of episodes and composition of time scales. The applicability on real world NDVI time series was demonstrated by mapping the coupling between precipitation variability, interannual vegetation changes, and the El Niño Southern Oscillation and Indian Ocean Dipole phenomena for ecoregions in East and Central Africa. In most areas where precipitation was found sensitive to oceanic forcing, the EEMD detected vegetation changes matched the predicted response, except in dense forest ecosystems.**

Keywords: Time series analysis, Vegetation monitoring, Interannual variability, Ensemble Empirical Mode Decomposition, NDVI, Longitudinal data

## 1. Introduction

Interannual changes in vegetation cover and productivity are key to understanding the impacts of global climate change on ecosystem dynamics (Hilker et al. 2014; Luo et al. 2011). Such year to year variations in vegetation status critically influence ecosystem services such as terrestrial carbon sequestration (Piao et al. 2011) and regulation of the hydrological cycle (Liu et al. 2008). However, the response of vegetation to climatic anomalies is often ambiguous (Brando et al. 2010), variable across ecosystems (Holmgren et al. 2013; Piao et al. 2014) and subject to critical transitions at tipping points (Hirota et al. 2011). Therefore its quantification requires dedicated monitoring tools.

Satellite remote sensing is an excellent tool to assess land ecosystem dynamics as it provides periodic, spatially explicit and consistent measurements of biophysical processes at the Earth's surface (Baynard 2013). It has become an established monitoring instrument in the fields of forestry (Achard et al. 2002), agriculture (Atzberger 2013) and plant ecology (Xie et al. 2008). Several vegetation indices and biophysical parameter products have been derived over the past 35 years from different satellite sensors, e.g., NOAA-AVHRR (Cracknell 2001), SPOT-VGT (Maisongrande et al. 2004), MODIS (Huete et al. 2002) and PROBA-V (Dierckx et al. 2014). A key product in these datasets is the Normalized Difference Vegetation Index (NDVI) which highlights vegetation greenness and photosynthetic activity (Tucker and Sellers 1986). The combined data archives form a unique long term NDVI record which has been used in numerous studies dealing with the effects of climate variability on vegetation status in space and time (e.g., Fensholt et al. 2012; Guo et al. 2014).

This paper proposes and validates a framework for a tool to detect interannual NDVI fluctuations, by considering the ensemble of past efforts in remote sensing data processing and introducing promising advances from another domain of science, i.e., signal processing.

Early approaches accounted for interannual variability in time series *implicitly* without explicitly separating the interannual component. In pioneer works (Anyamba and Eastman 1996; Nicholson

and Farrar 1994; Richard and Poccard 1998), the forcing role of climate variability on NDVI was explored by spatial representations of the (lagged) correlation between mean monthly or annual NDVI values and corresponding aggregate rainfall values. This principle of lagged linear response was further explored in a simple linear regression approach (Camberlin et al. 2007; Zhou et al. 2009; Zhou et al. 2014a), whereas other authors (De Keersmaecker et al. 2015; Myneni et al. 1996; Plisnier et al. 2000) introduced the use of standardized anomalies from the average annual season as the key to reveal interannual variability. Brown *et al*. (2010) elaborated this notion of climatic forcing at multiple interannual time scales, supported by an increasing insight in the mechanisms and the effects of regional climate oscillations (Holmgren et al. 2001; Nicholson and Kim 1997; Williams and Hanan 2011).

Other tools have been developed to *explicitly* separate different time scale components in NDVI time series using techniques from the signal processing domain. Methods range from Fourier spectral analysis (Immerzeel et al. 2005; Lhermitte et al. 2008; Roerink et al. 2003) and wavelet analysis (Torrence and Compo 1998) to specific tools to detect, separate and model intra-annual variability, long term trends and abrupt changes from satellite image time series: TIMESAT (Jönsson and Eklundh 2002), BFAST (Verbesselt et al. 2010) and SPIRITS (Eerens et al. 2014).

Despite this wide range of tools, there is no quantitative method that focuses on interannual fluctuations, which we believe are crucial in understanding and predicting the response of ecosystems to an increasingly variable climate. For example, Fourier analysis methods generally focus on mapping the occurrence of annual unimodal or bimodal growing seasons to distinguish land use patterns (e.g., Lhermitte et al. 2011), while longer time scales are only represented as harmonic frequencies which are not variable over time. Wavelet analysis provides a more flexible technique to detect interannual components which are themselves variable over time, beyond the static and strictly harmonic Fourier framework. It has therefore been applied in numerous studies to identify interannual components in NDVI and climatic time series (Galford et al. 2008; Martínez and Gilabert

2009; Quiroz et al. 2011). Swinnen (2008) used wavelet coherency, a form of cross-spectral analysis (Torrence and Webster 1999) to systematically examine the climate-vegetation coupling over all occurring time scales in the time-frequency domain.

However, a generic tool to study interannual NDVI components must yield easily interpretable components in the time domain, so as to not impose constraints for further processing (e.g., co-analysis with climatic series), and to allow integrated use with other time series toolboxes. An answer to the need of breaking down an NDVI series into different time scale components while preserving the time domain flexibility of the wavelet approach is found in the Empirical Mode Decomposition (EMD) methodology, proposed and elaborated by Huang and colleagues in their key publication (Huang et al. 1998). The EMD algorithm (summarized in Section 2.2, and detailed in Appendix A) iteratively extracts the intrinsic time scales from the series, yielding a finite set of components with decreasing frequency and a residual trend component. EMD has been applied to climatic or biogeophysical time series to analyze climate variability in general (Coughlin and Tung 2005; Molla et al. 2011; Pegram et al. 2008), to model slow components in climate simulations (Brisson et al. 2015), to assess climatic effects on plant phenology (Guan 2014), to aid crop classification (Chen et al. 2011) and to remove the effect of platform orbital drifts in cross-sensor image time series (Pinzon et al. 2005).

EMD and its extension Ensemble Empirical Mode Decomposition (EEMD; Wu and Huang 2009) are straightforward as a decomposition algorithm, except for the procedure to sort ensembles of components lacking explicitness, which we tackle in this paper. However, processing EEMD components to robust indicators of interannual variability has not been addressed yet and faces a number of problems. First, the capability of assigning components to their underlying physical processes (e.g., noise, annual climatology) has been demonstrated *ad hoc* (Coughlin and Tung 2005), but has not been formalized to make it applicable for systematic processing of large spatiotemporal datasets. Second, distinguishing significant signal components from noise-induced components

105 critically determines the method's performance and the conclusions that can be drawn from its
106 output. Significance tests for (E)EMD components exist but lack validation, which is a prerequisite
107 for a generic processing tool.

108 The overall aim of this work is to introduce the EEMD technique in remotely sensed image
109 processing applications, and to propose a framework for its implementation and validation. More
110 specifically, the objectives of this paper are:

111 - To define a robust procedure based on EEMD to separate interannual components in
112 remotely sensed NDVI time series from annual and noise components;
113 - To assess the performance of this procedure and its sensitivity to EEMD model
114 parameters and characteristics of the input time series;
115 - To demonstrate its applicability in a case study over East and Central Africa.

116 Based on the hypothesis that climate variability is reflected in NDVI time series as relatively
117 weak, irregular and slow fluctuations, we composed simulated time series containing annual,
118 interannual and noise components. In this way, we set up a validation framework in which the
119 interannual component detected after decomposition can be compared to the original input
120 component. After this validation, we assess the representativeness of our simulations for real world
121 situations by examining the detected interannual NDVI response to precipitation in East and Central
122 Africa. In the areas where precipitation is sensitive to regional climate variability, a corresponding
123 vegetation response is expected in ecosystems with limiting precipitation.

## 124 2. Data and methods

125 A remotely sensed NDVI dataset spanning the 1981-2014 period and covering a wide range of
126 climate types and ecosystems (East and Central Africa) along with series of historic climate indices
127 were used to set up a framework to test and validate the envisaged tool for interannual component
128 detection. We adopted the Ensemble Empirical Mode Decomposition (EEMD) to obtain a per-pixel

129 estimate of the constituent components of NDVI series. These estimates were used to simulate
130 annual, interannual and noise components with controlled variation of their parameters. A set of
131 mixtures of simulated components were decomposed using EEMD, and the interannual components
132 were reconstructed by three distinct approaches. Measures of correspondence between the input and
133 output interannual components were calculated to evaluate the performance and sensitivity to various
134 EEMD model parameters, data characteristics and reconstruction methods. Finally, the validated
135 method was applied on the original NDVI dataset and historic climate indices in a regional case
136 study.

137 ## 2.1. Data

138 ### 2.1.1. Remote sensing datasets

139 The Normalized Difference Vegetation Index (NDVI) is defined as the normalized difference in
140 surface reflectance ($Refl$) in the near-infrared ($NIR$) and red ($R$) wavebands of a sensor and is
141 considered to be a proxy for vegetation greenness (Tucker and Sellers 1986).

142 $$NDVI = (Refl_{NIR} - Refl_R)/(Refl_{NIR} + Refl_R) \quad \text{(Eq.1)}$$

143 Various efforts have been made to produce consistent long term NDVI records from sensor
144 sequences, each with their particular merits and flaws (Tian et al. 2015). Within the setup of this
145 paper, residual inconsistencies in the baseline dataset do not fundamentally affect the data
146 simulations or the validation results produced from them. Therefore, this section is limited to a
147 description of the underlying datasets whereas the implications of data inconsistencies are discussed
148 at length in the discussion section of this paper.

149 NASA's Long Term Data Record (LTDR) contains the reprocessed daily global AVHRR
150 imagery from subsequent NOAA platforms (1981-1999 in version 2; extended with 2001-2013 in
151 version 4), at 0.05° resolution. The LTDR reprocessing aims to achieve data consistency through
152 both improved radiometric calibration (Vermote and Kaufman 1995), corrections for orbital drift

153 effects (Nagol et al. 2014; Vermote et al. 2009) and atmospheric corrections with ancillary
154 meteorological data (Pedelty et al. 2007).

155 The SPOT-VEGETATION (VGT) programme (1998-2014) delivered global NDVI images of 1
156 km resolution as ten-daily maximum NDVI synthesis products as to minimize cloud effects present
157 in daily recordings (Deronde et al. 2014). Residual clouded pixels typically cause sharp troughs in
158 temporal NDVI profiles. These are detected and their values replaced by means of an NDVI
159 interpolation algorithm (Eerens et al. 2014; Swets et al. 1999) to represent a more realistic smooth
160 temporal NDVI profile.

161 One approach to obtain a long uninterrupted baseline dataset is to merge the VGT archive (1998-
162 2014) with the first part of the LTDR archive (1981-1999). Both image datasets were resampled to a
163 common systematic 20 km frame and subjected to the same synthesis and smoothing procedures to
164 yield comparable NDVI series. Finally, cross-sensor calibration reduces the inconsistencies between
165 NDVI from the AVHRR and VGT sensors caused by differences in sensor spectral responses
166 (Trishchenko et al. 2002), native spatial resolutions (Tarnavsky et al. 2008), platform orbits and
167 sensor geometries (Swinnen and Veroustraete 2008). Empirical cross-sensor calibration was
168 achieved following the linear VGT-to-AVHRR correction model proposed by Steven et al. (2003).
169 We re-estimated the correction equation from corresponding pixels in images from 1999, applying
170 geometric mean regression (Ji and Gallo 2006).

171 $$NDVI_{AVHRR} = 0.799\, NDVI_{VGT} + 0.081 \; (RMSE = 0.069\, NDVI) \qquad \text{(Eq.2)}$$

172 2.1.2. Climate data

173 Interannual climatic variability is represented in this study by two climatological indices which
174 have been linked to interannual rainfall patterns in East and Central Africa (IPCC 2007). The
175 Oceanic Niño Index (ONI) (NOAA-CPC 2014) represents sea surface temperature anomalies in the
176 Pacific Ocean (ERSST v4; Huang et al. 2014) and is the most widely used indicator for the episodes

of the El Niño Southern Oscillation (ENSO), which are known to have impacts globally. In addition, East Africa's rainfall is largely driven by the Indian Ocean dynamics, which as well display interannual episodes as summarized in the Indian Dipole Mode Index (DMI) (JAMSTEC 2010). This is the difference in sea surface temperature anomalies between the western and eastern equatorial Indian Ocean. Monthly values for both indices were smoothed with a 12-month running window in order to obtain an estimate of the interannual climatic forcing (Figure 1).

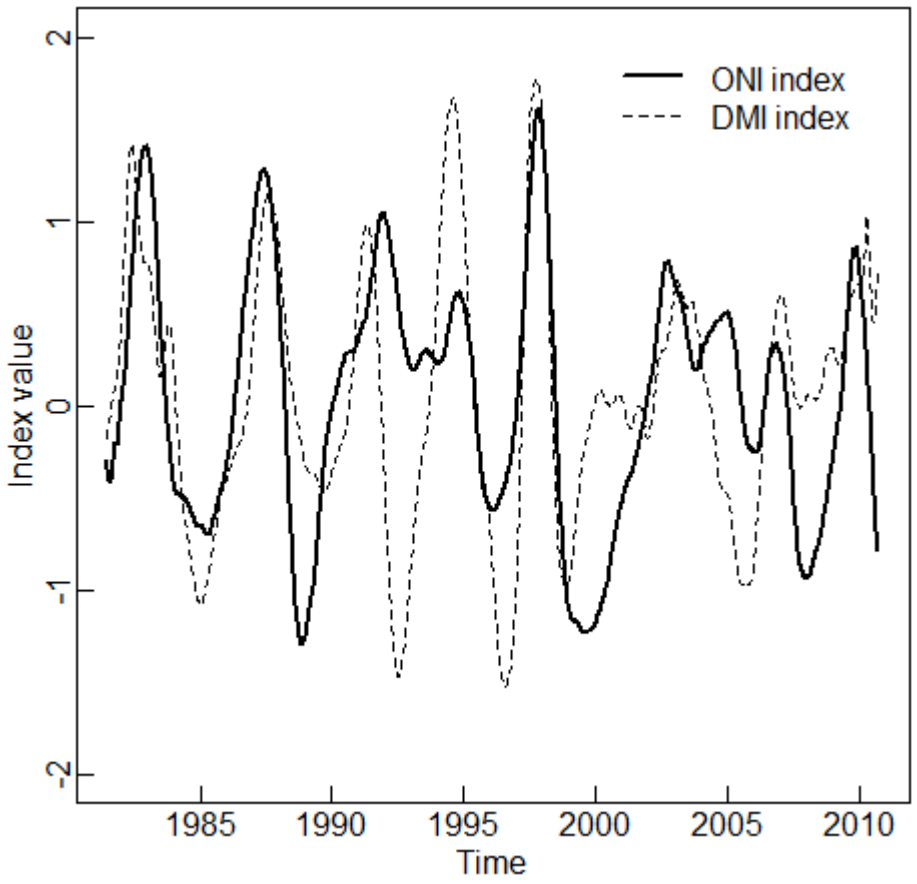



*Figure 1. The Oceanic Niño Index (ONI) and the Dipole Mode Index (DMI) are derived from sea temperature measurements and represent the positive and negative episodes of climate oscillations that affect interannual rainfall patterns over East and Central Africa (IPCC 2007).*

Gridded precipitation estimates derived from satellite observations (1983-2012) are obtained from the PERSIANN system (Sorooshian et al. 2000). The PERSIANN system merges infrared

189 brightness images with rainfall estimates from satellite microwave data to daily 0.25° precipitation
190 maps, with error detection and quality control using ground observations (Hsu et al. 1999).

191 2.2. Ensemble EMD decomposition model

192 For a detailed description of the basic EMD model, we refer to Huang *et al*. (1998), whereas the
193 algorithmic implementation and its parameters are described in Appendix A. Briefly, EMD works by
194 connecting local maxima (minima) in a series X(t) by a cubic spline which yields upper (lower)
195 envelopes. The time series minus the mean of the upper and lower envelopes contains the high
196 frequency information. In a process termed *sifting*, the above steps are repeated to remove residual
197 low frequency information until the number of extrema and zero-crossings does not change for *S*
198 consecutive siftings. The outcome is a component of the original series that isolates a specific
199 frequency of the series, termed an Intrinsic Mode Function (IMF). The residual of the data and a
200 detected IMF is again subjected to the sifting process to iteratively yield a set of $k$ IMFs ($IMF_{i=1\ldots k}$)
201 with decreasing frequency, until a final residual term R(t) has less than two local extrema.

202 $$X(t) = \sum_i IMF_i(t) + R(t) \qquad \text{(Eq.3)}$$

203 Although powerful, basic EMD often leads to one time scale intermittently partitioned over two
204 or more IMFs, an issue referred to as *mode mixing* (Huang et al. 1999), illustrated in Figure 2a. To
205 cope with mode mixing, Wu and Huang (2009) proposed Ensemble EMD (EEMD) as an extension
206 to the method. By iteratively adding a finite amount ($\sigma_{add}$) of Gaussian white noise to the series,
207 applying EMD to the mixture, grouping corresponding IMFs into ensembles, and averaging each
208 ensemble of IMFs over a large number $N_n$ of noise realizations, mode mixing is largely eliminated
209 (Figure 2b).

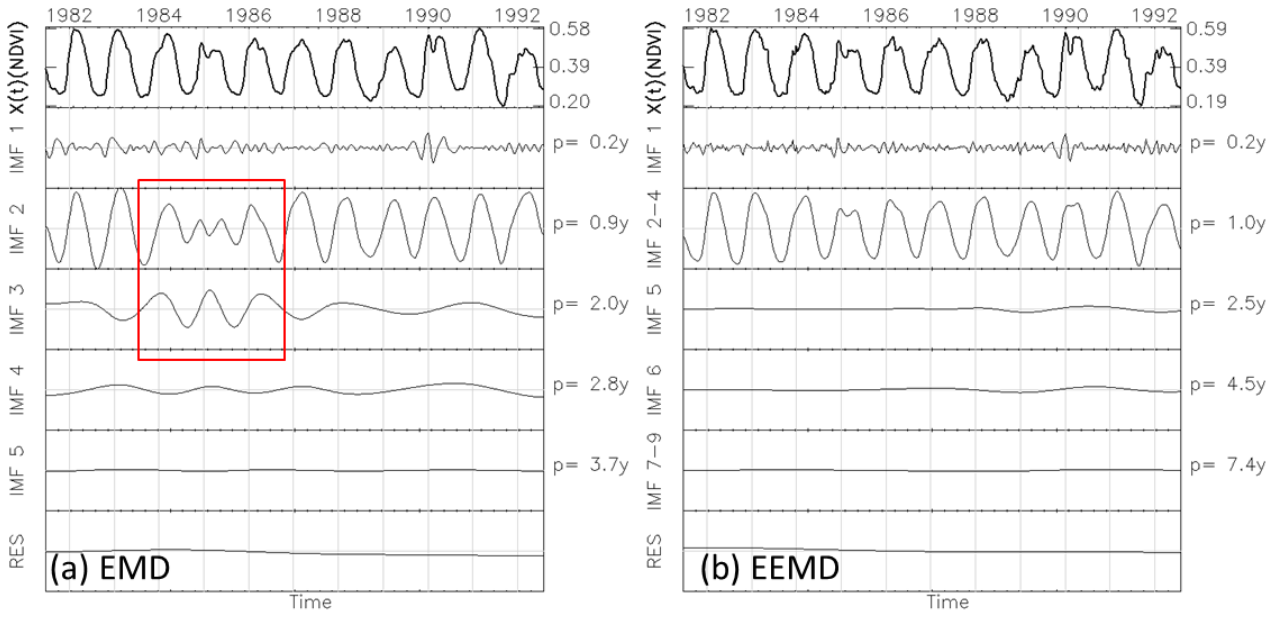



*Figure 2. A temporal NDVI profile (pixel 7 in Figure 3, North Tanzania) and its Intrinsic Mode Functions by (a) EMD decomposition and (b) the improved EEMD decomposition. EEMD decreases 'mode mixing' whereby fragments of a single time scale occur in another component, e.g. for the annual time scale for the period 1984-1986 (box on graph (a)).*

Although EEMD was applied recently by various authors (e.g., Feng et al. 2012; Guan 2014; Kuo et al. 2013; Zhou et al. 2014b), no formal definition of *corresponding* IMFs, required to sort them into ensembles, has been given. However, automated EEMD analysis of a large array of per-pixel NDVI time series requires a robust sorting procedure, certainly because different noise realizations can yield slightly different numbers of IMFs, and assigning each IMF to its appropriate group (or *bin*) before averaging is crucial in attaining successful mode separation. To formalize the IMF sorting step, we propose to use IMF periods $p_i$ (i.e., the series length divided by twice the number of zero-crossings in $IMF_i$) as the basic criterion, where the EMD decomposition of the original series without added noise defines reference periods $p_{ref.1},\ldots,p_{ref.k}$. Since EMD is found to produce components with a general pattern of doubling (or exponentially growing) periods (Flandrin et al.

2004; Wu and Huang 2004), the geometric means of subsequent reference periods $p_{ref.j}$ are considered as the bin boundaries for assigning IMFs to a bin after each noise realization.

$$bin_j = \left[\sqrt{p_{ref.j-1}p_{ref.j}}\,,\sqrt{p_{ref.j}p_{ref.j+1}}\right] \quad \text{(Eq.4)}$$

The sorting algorithm thus evaluates the IMFs of each noise-distorted ensemble member, and assigns them to their proper pre-defined period bins. The IMFs in a bin are then averaged over the number of noise realizations $N_n$, to yield a decomposition with better mode separation. Our implementation uses parameter values for added noise amplitude [$\sigma_{add} = 0.1$] as in other EEMD applications (Kuo et al. 2013; Wu and Huang 2009) and [$N_n = 50$] since the slight differences in outcomes for higher $N_n$ did not offset linearly increasing processing time.

## 2.3. Simulated time series

Our central hypothesis is that observed NDVI time series are composed of an annual component encompassing one or two growing seasons, an interannual component including climatic influence as well as residual effects of sensor merging and orbital drift, and a high-frequent noise component reflecting residual sensor noise, radiometric and geometric errors (Lhermitte et al. 2011). Each of these components can be present in varying amplitude levels and under varying degrees of irregularity in timing. For the sensitivity analysis of the EEMD model's performance to be representative and the detection method to be generic, these forms of variability must be parameterized and the parameter ranges be estimated from a real NDVI dataset, here the intercalibrated AVHRR/VGT dataset described in Section 2.1.1.

### 2.3.1. Study area stratification

Figure 3 shows the study area represented by each pixel's mean NDVI value ($\mu_{tot}$) over the entire series, thereby revealing the major ecoregions in East and Central Africa. The series of 9 pixels with distinct NDVI levels and seasonal profiles were selected across climatic regions to

optimally explore parameter ranges related to NDVI data. Figure 4 depicts cross-calibrated temporal NDVI profiles derived from the combined AVHRR and VGT datasets for this set of pixels.

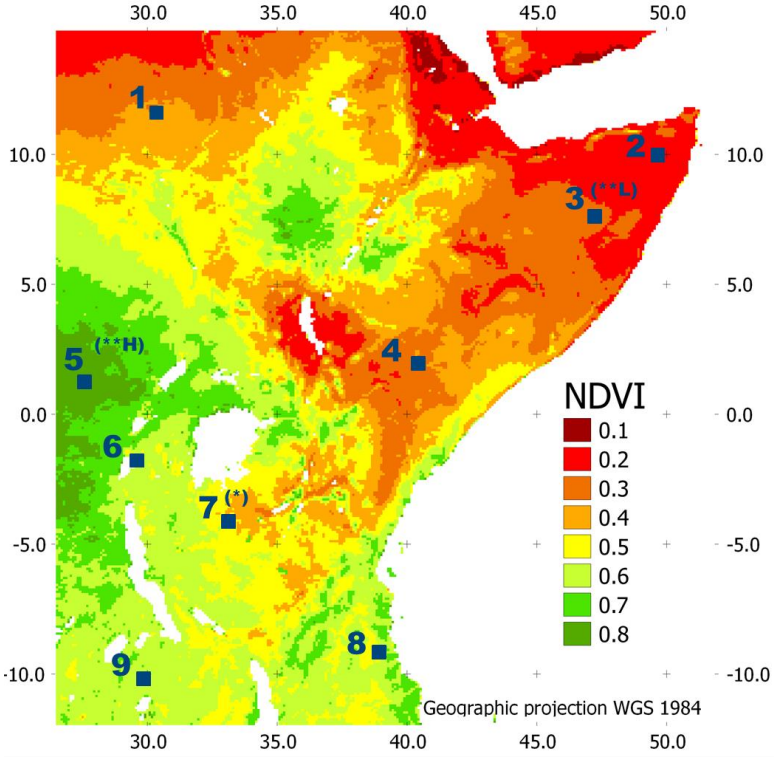



*Figure 3. Mean NDVI value over the period 1981-2012 per image pixel over East and Central Africa. The square symbols indicate the locations of the selected pixels, used to estimate parameters for NDVI time series simulation (Table 1).*

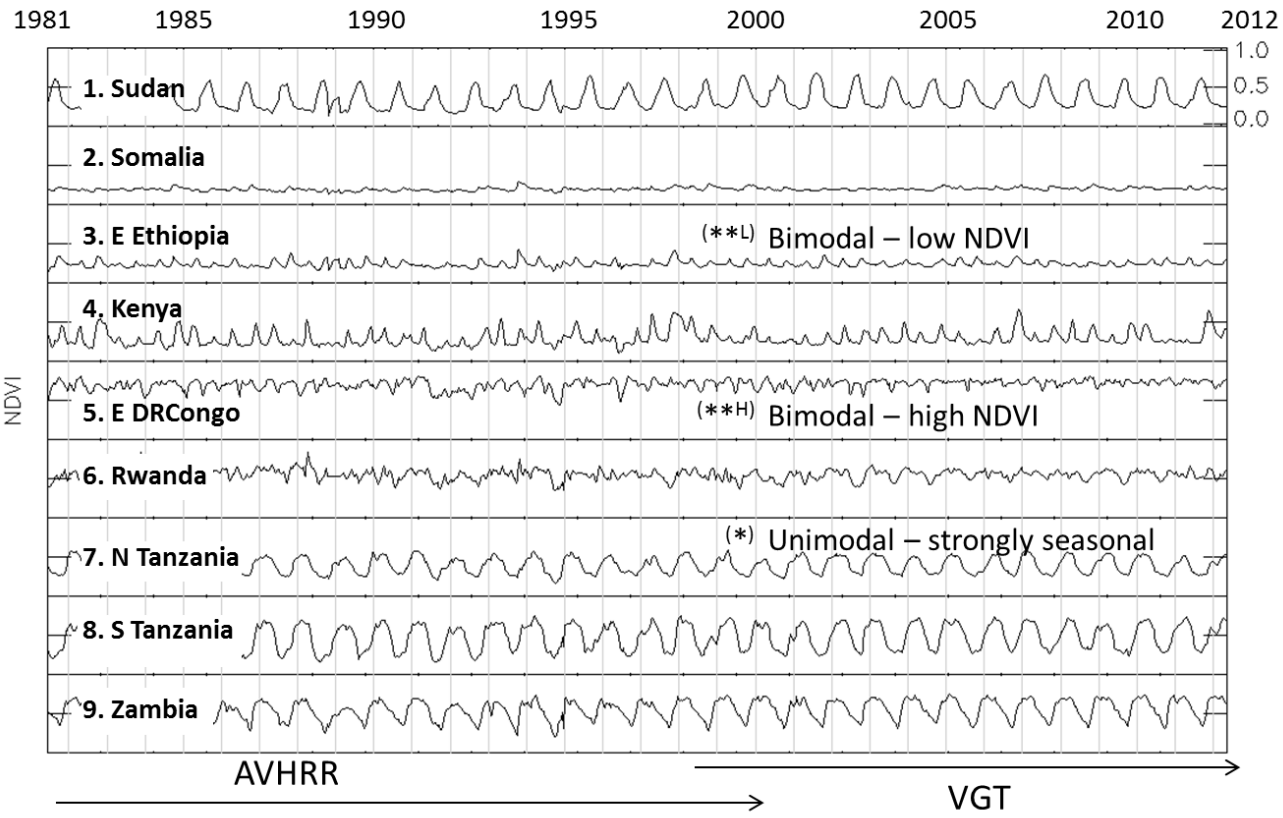



*Figure 4. Time series of the Normalized Difference Vegetation Index (NDVI) for 9 pixels in distinct climate zones in East and Central Africa (Figure 3), derived from cross-calibrated image data from the NOAA-AVHRR (1981-1999) and SPOT-VGT (1998-2012) satellite sensors. Three prototypes of growing seasons were selected for data simulations: (*)unimodal (one rainy season) and (**)bimodal (two rainy seasons) with either overall low (L) or high (H) NDVI levels.*

### 2.3.2. Estimation of variability in annual, interannual and noise components

Without any prior knowledge of relative signal amplitudes in real NDVI time series, a first estimate was made from their respective IMFs. The IMFs resulting from EEMD of each of the nine series were grouped into noise IMFs, annual IMFs and interannual IMFs, following the bin boundary logic as described in Section 2.2. Considering the occurrence of bimodal growing seasons, 0.5 year and 1 year are taken as the theoretical periods for annual cycles. As IMFs normally fall within period bins of doubling frequencies, the bin boundaries are set at $0.5/\sqrt{2}$ years and $1.0\sqrt{2}$ years. Shorter

period IMFs are labeled as noise, whereas the sum of longer period IMFs is a first representation of the interannual component. The amplitude of each component is calculated as its standard deviation ($\sigma$). Furthermore, mean NDVI values ($\mu_{tot}$) and the numbers of modes in the average yearly growing season were tabulated per series (Table 1). Three prototypes of NDVI seasonality were identified for further modeling: a weak bimodal season with high NDVI $^{(**H)}$ (humid conditions, e.g., East DR Congo), strongly seasonal unimodal season $^{(*)}$ (subhumid conditions, e.g., North Tanzania), and an erratic bimodal season with low NDVI $^{(**L)}$ (semi-arid conditions, e.g., East Ethiopia).

*Table 1. Estimated parameters for the time series of 9 selected pixels: amplitudes of noise ($\sigma_{noise}$), annual ($\sigma_{ann}$) and interannual ($\sigma_{int}$) components, total amplitude ($\sigma_{tot}$) and mean value ($\mu_{tot}$) of the series, and the number of modes in the annual growing season. Three prototype seasons for further modelling are labeled as unimodal $^{(*)}$ with strong seasonality and bimodal $^{(**)}$ with low $^{(L)}$ and high $^{(H)}$ NDVI.*

| ***Location*** | $\sigma_{tot}$ | $\sigma_{noise}$ | $\sigma_{ann}$ | $\sigma_{int}$ | $\sigma_{int}/\sigma_{ann}$ | $\mu_{tot}$ | # modes |
|---|---|---|---|---|---|---|---|
| *1. Sudan* | 0.14 | 0.02 | 0.12 | 0.03 | 0.27 | 0.35 | 1 |
| *2. Somalia* | 0.02 | 0.01 | 0.01 | 0.02 | 1.09 | 0.20 | 2 |
| *3. E Ethiopia* $^{(**L)}$ | 0.04 | 0.02 | 0.03 | 0.02 | 0.66 | 0.24 | 2 |
| *4. Kenya* | 0.09 | 0.01 | 0.07 | 0.05 | 0.64 | 0.29 | 2 |
| *5. E DRCongo* $^{(**H)}$ | 0.06 | 0.03 | 0.04 | 0.03 | 0.74 | 0.70 | 2 |
| *6. Rwanda* | 0.07 | 0.03 | 0.04 | 0.03 | 0.69 | 0.54 | 2 |
| *7. N Tanzania* $^{(*)}$ | 0.10 | 0.01 | 0.10 | 0.02 | 0.19 | 0.39 | 1 |
| *8. S Tanzania* | 0.17 | 0.02 | 0.15 | 0.03 | 0.22 | 0.47 | 1 |
| *9. Zambia* | 0.12 | 0.03 | 0.11 | 0.03 | 0.23 | 0.53 | 1 |

#### 2.3.3. Annual component

Growing seasons generally appear as peaks in NDVI time series, with a certain degree of irregularity. A flexible parametric model for a seasonal NDVI peak was adopted from Jönsson and Eklundh (2002). Its basic approach is to fit an asymmetric Gaussian model function $f(t)$, defined by its base level b, peak position p and amplitude A, and the width $w_{l,r}$ and flatness $f_{l,r}$ of both tails (see

Equations 5 and 6). This approach was expanded to the case of bimodal growing seasons which are modeled as two additive peaks with different peak locations within the year. A unimodal or bimodal peak was thus fitted to each of the 31 growing seasons of the three prototype series spanning 1981 to 2012. Sample means and standard deviations (n = 31) give an account for the expected values and natural variation of the seasonal parameters throughout the study area (Table 2). These sample statistics were used to generate simulated annual components as a function of time $f(t)$ with realistic shapes (base level, amplitude, peak widths and flatness) and controlled variation in irregularity (peak positions).

$$f(t) = b + A.g(t) \quad \text{(Eq.5)}$$

$$with\ g(t) = \begin{cases} exp[-((t-p)/w_l)^{f_l}];\ if\ t > p \\ exp[-((t-p)/w_r)^{f_r}];\ if\ t < p \end{cases} \quad \text{(Eq.6)}$$

*Table 2. Sample statistics from time series of 3 prototype growing seasons, denoted as mean (standard deviation). Five parameters represent the shape and timing of the first (I) and second (II) growing season: base level* b, *peak position* p *and amplitude* A, *and the width* $w_{l,r}$ *and flatness* $f_{l,r}$ *of the tails (as in Jönsson and Eklundh 2002; see Equations 5 and 6).*

| | b *(NDVI)* | A (NDVI) | | p (years) | | $W_l$(years) | | $W_r$(years) | | $f_l$ | | $f_r$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Seasonality type*** | | I | II | I | II | I | II | I | II | I | II | I | II |
| ***$^{L}$ Bimodal (Low NDVI)*** | 0.17 (0.04) | 0.25 (0.09) | 0.30 (0.10) | 0.37 (0.03) | 0.97 (0.04) | 0.15 (0.12) | 0.14 (0.11) | 0.10 (0.03) | 0.07 (0.05) | 1.39 (0.85) | 1.86 (1.50) | 1.62 (0.62) | 2.19 (1.78) |
| ***$^{H}$ Bimodal (High NDVI)*** | 0.61 (0.09) | 0.21 (0.07) | 0.19 (0.05) | 0.38 (0.09) | 0.83 (0.09) | 0.30 (0.17) | 0.23 (0.17) | 0.15 (0.09) | 0.17 (0.11) | 2.32 (1.64) | 2.88 (2.53) | 2.16 (1.32) | 2.46 (1.70) |
| ** Unimodal (Strongly seasonal)* | 0.23 (0.03) | 0.32 (0.05) | - | 0.45 (0.10) | - | 0.29 (0.10) | - | 0.28 (0.09) | - | 2.64 (1.42) | - | 2.51 (1.10) | - |

For each of the seasonality types, sequences of growing seasons were generated with their respective mean values of base level, amplitude and peak positions, widths and flatness. Seasonal irregularity ($\varphi_{seas}$) was introduced by adding a random variation of 0.0, 0.5, 1.0 and 1.5 times the standard deviations for the respective peak positions.

306 2.3.4. Interannual component

307 Realistic and representative a priori simulations of interannual variability are more difficult to
308 obtain, since they are the outcome our tool aims to deliver. Therefore we combine descriptions found
309 in literature on the impact of interannual climate variability on vegetation with our additive
310 hypothesis and with the estimated levels of variability from the model series set of our study area.
311 The major coupling between climatic variability and vegetation phenology is thought to act through
312 episodes of severely increased or decreased precipitation (Holmgren et al. 2013). One well-known
313 cause of such episodes is the El Niño Southern Oscillation (ENSO), a recurrent climate phenomenon
314 at a 3 to 7-yearly time scale (IPCC 2007) which is known to force vegetation growth beyond its
315 seasonal cycle (Holmgren et al. 2013). Although the interactions are complex in nature and region
316 dependent (Sitters et al. 2012), increased precipitation episodes generally cause a rapid, but lasting
317 response over several years (Holmgren et al. 2001).

318 Within the limitations of our additive model, we thus represent climate-driven interannual
319 variability in NDVI as a weak modulation, with a steep increase and a relative slow decline. We
320 employ again the asymmetric Gaussian simulation model to control the corresponding parameters.
321 ENSO peak positions ($t_{ONI}$) and peak-to-peak amplitude variations ($A_{ONI}$) were derived from the
322 historic ONI series, to reconstruct historic ENSO episodes and estimate the natural variability in
323 duration of episodes ($\mu_{t.ONI}$; $\sigma_{t.ONI}$) for the simulation of irregular episodes with relative magnitudes
324 ranging from 1 to 3 ($R_{1\text{-}3}$). Table 3 lists the resulting set of interannual models. The overall
325 interannual amplitude levels ($A_{int}$) were defined as fractions of the annual amplitude ($\sigma_{int}/\sigma_{ann}$),
326 which were in turn estimated from the interannual IMFs obtained by EEMD decomposition of the
327 NDVI model series set. Based on the range of estimated annual/interannual ratios in Table 1, this
328 parameter takes on values 0.1, 0.2, 0.33, 0.5 and 1.0 in the simulations.

329

*Table 3. Various models for interannual variability were based on the historic ONI index, defined by the timing of its episodes ($t_{ONI}$) and their amplitudes ($A_{ONI}$) with optional random fluctuations at three different levels $R_{1-3}$. The sample statistics of episode duration ($\mu_{t.ONI}, \sigma_{t.ONI}$) are used to model irregular episodes (*I*).*

| Label | Description | Parameters |
|---|---|---|
| $H_{oni}$ | Historic Oceanic Niño Index | $(t_{ONI}, A_{ONI}) \rightarrow \mu_{t.ONI}, \sigma_{t.ONI}$ |
| $H_{cst}$ | Historic episodes, constant amplitude | $f(t_{ONI}, A_{int})$ |
| $H_{var}$ | Historic episodes, historic varying amplitude | $f(t_{ONI}, A_{int} * A_{ONI})$ |
| $R_{cst}$ | Regular episodes, constant amplitude | $f(\mu_{t.ONI}, A_{int})$ |
| $R_{rnd}$ | Regular episodes, random varying amplitude | $f(\mu_{t.ONI}, A_{int} * R_{1-3})$ |
| $I_{cst}$ | Irregular episodes, constant amplitude | $f(\mu_{t.ONI} + \sigma_{t.ONI}, A_{int})$ |
| $I_{rnd}$ | Irregular episodes, random varying amplitude | $f(\mu_{t.ONI} + \sigma_{t.ONI}, A_{int} * R_{1-3})$ |

### 2.3.5. Noise component

Real world NDVI noise levels ($\sigma_{noise}$) were estimated as the root mean square (*rms*) amplitudes of the sum of high frequency IMFs of each of the model set series (Table 1), as these IMFs are assumed to contain nearly all noise fragments. In order to assess the sensitivity of EEMD performance to noise in the input time series, we added random noise with *rms* amplitudes 0.00, 0.01, 0.02, 0.03 and 0.05 NDVI. The noise model present in real NDVI series is unknown a priori. As Mann and Lees (1996) argue, geophysical phenomena display a certain degree of persistence of random distortions, quantified as the noise lag-1 autocorrelation ($AC_1$) or *redness*. Therefore we employed both white noise ($AC_1 = 0$) and red noise ($AC_1 = \{0.25, 0.50\}$) simulations drawn from a normal distribution.

The full set of simulated series was generated by adding each combination of annual, interannual and noise component, varying over 3 seasonality types, 4 levels of seasonal irregularity ($\varphi_{seas}$), 7 interannual models, 5 interannual to annual amplitude ratios ($\sigma_{int}/\sigma_{ann}$), 3 noise models ($AC_1$) and 5 noise levels ($\sigma_{noise}$). Sensitivity of the method's performance to reconstruct interannual components

was not only tested for data-related variations, but also for a method-related parameter, as described in the next paragraph.

### 2.4. Components post-processing

The general approach to reconstructing the interannual component from EEMD components is to add combinations of detected longer-than-annual components. Estimation of components' periods based on the number of zero-crossings was outlined above. A currently unresolved issue in the post-processing approach is the selection of *significant* interannual components. Three different approaches to selection have been evaluated through sensitivity analysis.

The need for component selection originates from the potential occurrence of low frequency components in the decomposition of random noise present in the data. As discovered by Flandrin *et al*. (2004) and further examined by Wu and Huang (2004), (E)EMD breaks down white noise into components of approximately doubling periods. However, the energy (squared amplitude) of a white noise (E)EMD component is shown to be related to its period, and constrained by theoretical confidence bounds (see Appendix B). Monte Carlo verification of this relationship and its confidence bounds deliver a framework to distinguish a real signal component from noise-induced components, by checking whether its energy exceeds with high confidence the expected energy of a noise component with the same period. This framework is proposed by Wu and Huang (2005) and has been adopted for climatological (E)EMD applications (e.g., Coughlin and Tung 2005). The signal/noise distinction is however entirely dependent on the quantification of the noise level in the series. An 'a priori' test considers the entire series as potential noise, while an 'a posteriori' test can be performed by estimating the noise level from the high-frequent (E)EMD components which, under the additive hypothesis, are purely noise-induced (Wu and Huang 2005). However, we argue that a given (E)EMD component may simultaneously contain signal information and noise information with similar periods. Therefore, exclusion as well as inclusion will introduce some error

in the reconstructed interannual component. In case of sufficiently signal-to-noise ratios in the data, this balance may be in favor of retaining all detected components without significance test. We therefore consider 'no test', 'a priori' test and' a posteriori' test in the sensitivity analysis for the reconstruction of interannual components from simulated NDVI time series.

### 2.5. Performance measures

Successful detection and reconstruction of interannual components from a mix of known input components, and its sensitivity to data-related and model-related parameters must be expressed in quantitative measures. We propose three different performance measures to highlight different types of correspondence between two time series (Lhermitte et al. 2011), in this case the input $I_t$ and reconstructed output $O_t$ interannual components.

Correlation *r* between two time series indicates whether the detected pulses are correctly localized in time and have consistently the correct direction (positive or negative). It therefore indicates if statements on the occurrence in time of positive and negative climatic episodes are reliable.

$$r = \sum_t (I_t - \bar{I})(O_t - \bar{O}) \Big/ \left[ \sqrt{\sum_t (I_t - \bar{I})^2} \sqrt{\sum_t (O_t - \bar{O})^2} \right] \quad \text{(Eq.7)}$$

The relative root mean squared error (*rRMSE*) between two series summarizes the *rms* deviations between input and reconstructed components, relative to the interannual input component's *rms* amplitude. It indicates if statements on the magnitude of particular episodes are reliable.

$$rRMSE = \sqrt{\sum_t (I_t - O_t)^2 / \sum_t {I_t}^2} \quad \text{(Eq.8)}$$

Finally, weighted spectral coherence (*wCOH*) between two time series measures the correspondence (from 0 to 1) of the spectral content of two time series $X_t$ based on their Fourier spectra $F(X_t)$, weighted by the Fourier spectrum of the reference series (with $F^*$ the complex conjugate of F and $\langle . \rangle$ expectations estimated by ensemble averaging, see Alston et al. (2013) for a

detailed description). It therefore indicates the reliability of statements on the time scales of detected oscillations.

$$COH = |\langle F^*(I_t).F(O_t)\rangle|^2 \,/\, (\langle|F(I_t)|^2\rangle\langle|F(O_t)|^2\rangle) \quad \text{(Eq.9)}$$

$$wCOH = \sum(\mathrm{F}(I_t) * COH)/\sum F(I_t) \quad \text{(Eq.10)}$$

The overall strategy of composing a simulated series, its decomposition and subsequent evaluation of the detected interannual component against the input component is illustrated in Figure 5.

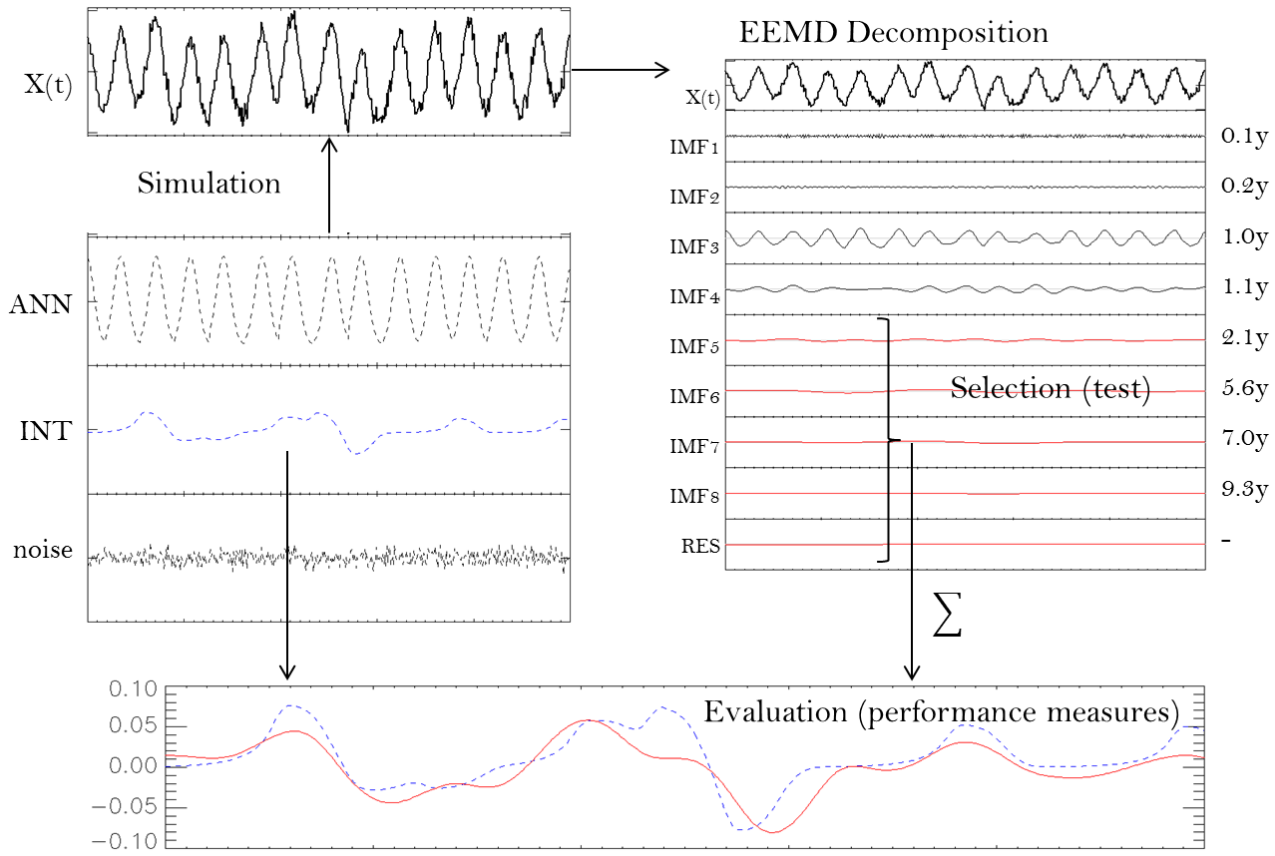



*Figure 5. Overview of the evaluation framework. Mixtures $X(t)$ of simulated annual (ANN), interannual (INT) and noise components are decomposed using EEMD. The interannual component (red solid line) is reconstructed from selected components and evaluated against the input component (blue dashed line).*

## 2.6. Sensitivity analysis

Estimating sensitivity of the above described performance measures to the parameters in the simulated series aims at identifying those factors which most affect the successful detection of interannual components. Interannual to annual amplitude ($\sigma_{int}/\sigma_{ann}$), seasonal irregularity ($\varphi_{seas}$), noise autocorrelation ($AC_1$), noise level ($\sigma_{noise}$), seasonality type (SEAS), interannual model (INT) and reconstruction method (METH) each partially explain the variance in performance, which is assessed through multiple linear regression. Real-valued variables ($\sigma_{int}/\sigma_{ann}, \varphi_{seas}, AC_1, \sigma_{noise}$) are standardized with respect to their respective sample means and standard deviations, in order to make the estimated regression coefficients comparable. Categorical variables (SEAS, INT, METH) enter the regression equation as dummy variables. The regression coefficient for each category represents the effect of the variable taking on that particular category with respect to a reference category, enabling a ranking of categories per performance measure. Within the approximation that an increase (decrease) of a real-valued variable by one standard deviation is equivalent to a categorical variable taking on a different category, all estimated coefficients are intercomparable and the factors that most affect the performance can be identified.

### 2.7. Case study: interannual NDVI and climate episodes

The validated performance of the EEMD method to retrieve interannual components from NDVI time series is conditioned by the degree to which the simulated series are representative for real world NDVI time series. To this end it was assessed in two steps whether the detected interannual NDVI patterns over East and Central Africa correspond to the expected long term fluctuations as predicted by historic climatic episodes.

In a first step, the zones where precipitation is most sensitive to interannual climatic and oceanic forcing are mapped by correlating the one-year accumulated Standardized Precipitation Index (SPI) with historic indices of two major climate phenomena acting in the region, i.e., ONI and DMI. The SPI index is  a z-score of a precipitation event compared to the historic sample of events in the same

period within the year, taking into account the non-normal distribution of rainfall events (Guttman 1999). The SPI algorithm allows the definition of a time scale, i.e., the width of the past time window to consider as an accumulated precipitation event. SPI with a one-year time scale was used as a first proxy of interannual precipitation. The resulting maps reveal where the climate phenomena described by the indices have a strong coupling with the interannual precipitation patterns. This is where a strong interannual NDVI response can be expected.

In a second step, the decomposition model is applied to the NDVI time series of each pixel in the AVHRR/VGT dataset over East and Central Africa. The detected interannual components per pixel (with the choice of model parameters guided by the outcomes of the sensitivity analysis) are in turn correlated with the ONI and DMI indices. If EEMD detection on NDVI time series is effective, the footprint of ONI and DMI must be found in the interannual vegetation changes in those areas sensitive to precipitation fluctuations revealed in step one. A portion of the Global Land Cover map (GLC2000; Bartholomé and Belward 2005) over the study area (Figure 6) assists in the interpretation of the vegetation responses to climate variability.

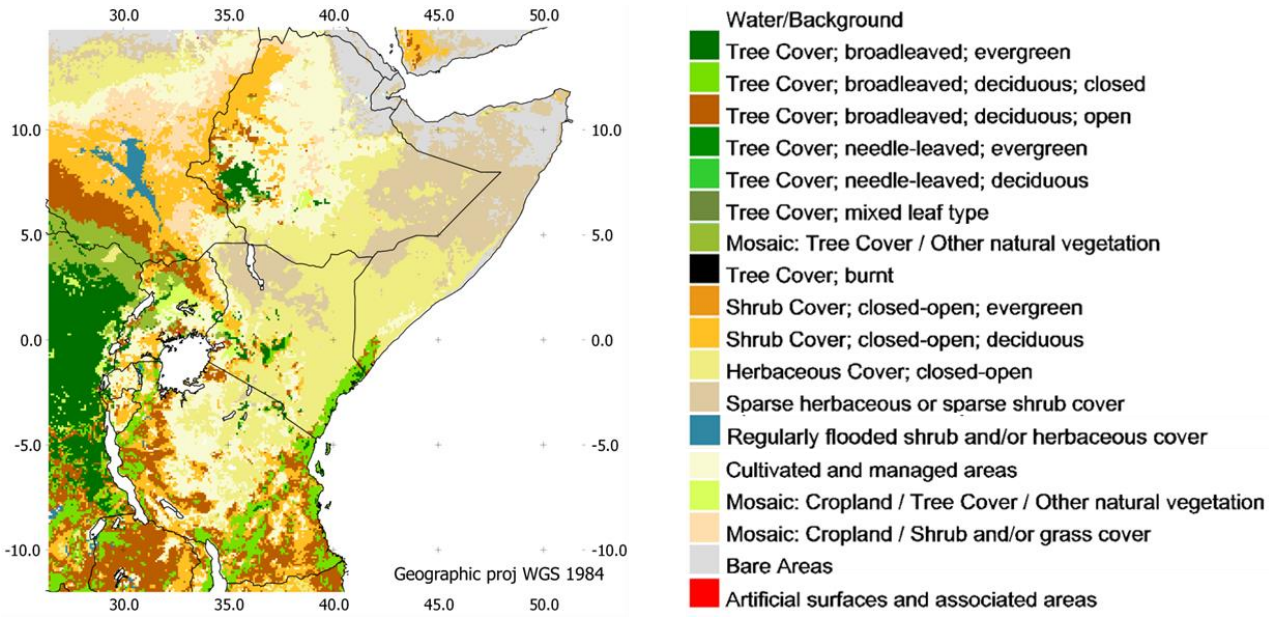



*Figure 6. The distribution of land cover types over the study area (GLC2000; Bartholomé and Belward 2005) reflects to a large extent the different levels of mean NDVI shown in Figure 3.*

*Ecosystems respond differently to regional precipitation variability, as confirmed in the demonstration case study over East and Central Africa (see Discussion section).*

## 3. Results

### 3.1. Sensitivity and performance of EEMD interannual NDVI detection using simulated time series

Following the evaluation framework (Figure 5), the sensitivity of the EEMD method was assessed through multiple regression of the performance measures against data and model parameters.

*Table 4. Multiple linear regression of 3 performance measures against data and model parameters indicates the sensitivities of the EEMD method. Real-valued variables are scaled in order to make their coefficients comparable to the dummy coefficients of categorical variables. Regression coefficients indicate the effect of a change of one standard deviation of a real-valued variable, or the effect of a change of category with respect to a reference category for categorical variables.*

| | **Correlation** | | **rRMSE** | | **wCOH** | |
|---|---|---|---|---|---|---|
| *Offset* | 0.68 | *** | 1.22 | *** | 0.41 | *** |
| *Real-valued variables* | | | | | | |
| $\sigma_{\mathrm{int}}/\sigma_{\mathrm{ann}}$ | 0.21 | *** | -0.41 | *** | 0.21 | *** |
| $\varphi_{seas}$ | -0.05 | *** | 0.30 | *** | -0.08 | *** |
| $AC_1$ | -0.08 | *** | -0.03 | *** | -0.06 | *** |
| $\sigma_{noise}$ | -0.01 | *** | 0.04 | *** | -0.02 | *** |
| *Categorical variables* | | | | | | |
| SEAS (reference category = 'high bimodal') | | | | | | |
| 'low bimodal' | 0.10 | *** | -0.42 | *** | 0.12 | *** |
| 'unimodal' | 0.05 | *** | -0.20 | *** | 0.12 | *** |
| INT (reference category = Hcst) | | | | | | |
| Rcst | 0.03 | *** | -0.07 | *** | 0.08 | *** |
| Rrnd | 0.00 | | 0.02 | | 0.05 | *** |
| Honi | 0.03 | *** | -0.03 | | 0.02 | *** |
| Icst | -0.02 | ** | 0.03 | . | -0.01 | |
| Hvar | -0.02 | *** | 0.02 | | -0.01 | |
| Irnd | -0.04 | *** | 0.02 | *** | -0.02 | ** |
| METH (reference category = 'no test' | | | | | | |
| 'a posteriori' | -0.16 | *** | 0.05 | *** | -0.09 | *** |

| | | | | | |
|---|---|---|---|---|---|
| 'a priori' | -0.34 | *** | -0.00 | -0.22 | *** |
| P(X>\|t\|) in Student-t: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | | |

Table 4 shows multiple linear regression coefficients of performance measures against parameters by least-squares estimation, along with their symbolized significance in a Student-t test.

Correlation between input and reconstructed output is primarily determined by the interannual/annual amplitude ratio ($\sigma_{int}/\sigma_{ann}$) and to the choice of post-processing method, where both 'a priori' and 'a posteriori' testing tend to reduce the correlation between input and output interannual components relative to the no-IMF selection method ('no test'). To a lesser degree, more autocorrelation in the noise model tends to distort correlations, whereas the overall noise level has little impact over the range estimated from the real world dataset. While there is a limited discrepancy in input-output correlation between the different seasonal models and its degree of irregularity, the choice of interannual model does not affect correlation strongly.

The rRMSE between input and output interannual components takes on values in a range around unity. It decreases significantly only in the case of a strong interannual component, a relatively regular annual cycle, or a particular type of growing season. Noise characteristics, interannual models and selection criteria for interannual IMFs have less weight on the detection accuracy measured by rRMSE.

Weighted spectral coherence (wCOH) shows sensitivities very similar to that of correlation, favoring the inclusion of all detected interannual IMFs in the estimated interannual component. Again, the relative strength of the interannual component compared to the seasonal cycle is of more importance than their assumed underlying models or the characteristics of the noise in the series.

As a conclusion from the correlation, rRMSE and spectral coherence measures (Table 4) it can be seen that the 'no test' approach performs best to detect the input interannual components, but that the performance of this method is still strongly dependent on the interannual/annual amplitude ratio and seasonal irregularity expressed as a number of sample standard deviations in start of season. This can

be seen for example in Table 5, which contains the values of the performance measures as a function of these two data-related parameters, following the 'no test' approach.

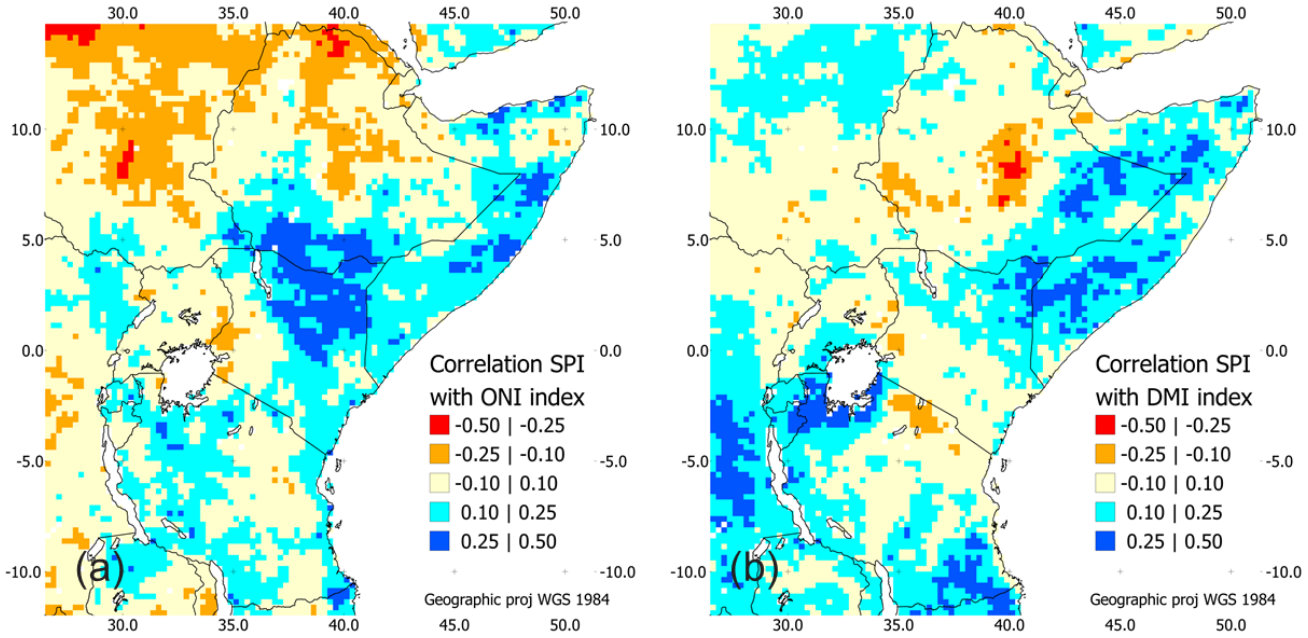
*Table 5. Performance measures of the 'no test' approach as a function of annual irregularity (ANN irr) and interannual/annual amplitude ratio (unimodal annual season, interannual model* $I_{rnd}$*, red noise (*AC1 *= 0.25;* $\sigma$ *= 0.03 NDVI)). Darker shades indicate better detection performance.*

| | | **Correlation** | | | | **rRMSE** | | | | **Weighted spectral coherence** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *ANN irr (σ)* | | *0.0* | *0.5* | *1.0* | *1.5* | *0.0* | *0.5* | *1.0* | *1.5* | *0.0* | *0.5* | *1.0* | *1.5* |
| *INT/ANN ratio* | *0.10* | 0.59 | 0.53 | 0.24 | 0.15 | 0.91 | 1.39 | 3.26 | 4.27 | 0.29 | 0.21 | 0.08 | 0.05 |
| | *0.20* | 0.86 | 0.70 | 0.41 | 0.49 | 0.54 | 0.76 | 1.69 | 1.36 | 0.65 | 0.28 | 0.15 | 0.23 |
| | *0.33* | 0.90 | 0.81 | 0.75 | 0.53 | 0.45 | 0.61 | 0.82 | 1.10 | 0.71 | 0.58 | 0.59 | 0.25 |
| | *0.50* | 0.91 | 0.88 | 0.82 | 0.65 | 0.45 | 0.49 | 0.66 | 0.90 | 0.70 | 0.70 | 0.55 | 0.31 |
| | *1.00* | 0.92 | 0.83 | 0.91 | 0.92 | 0.64 | 0.58 | 0.44 | 0.41 | 0.75 | 0.57 | 0.77 | 0.74 |

## 3.2. Case study over East and Central Africa

Figure 7 displays the correlation of precipitation patterns in East and Central Africa with dominant regional climate phenomena. The ONI index, representing the El Niño Southern Oscillation (ENSO) based on Pacific sea surface temperatures, explains a part of the precipitation variability principally in the regions most exposed to oceanic influence (Figure 7a). ENSO forcing is strongest in the Somalian, Ethiopian and Kenyan lowlands, only stretching further inland through the Turkana basin. Elsewhere its influence is limited by the Eastern and Southern rift mountains. Precipitation patterns in the Sudan lowlands towards the Sahel are rather out of phase with ENSO, pointing at the dominance of more local phenomena either not described by ENSO or negatively interfering with it (Williams and Hanan 2011). Plisnier et al. (2000) indicated that (a) the effect of global ENSO episodes is often overruled by local climate agents such as large lakes, (b) the Indian Ocean dynamics might provide a better proxy for climate variability in certain African regions. As seen from Figure 7b, The Indian Ocean DMI index, explains more variability of the Sahelian precipitation in the absence of direct ENSO influence. In addition, the precipitation downwind of

509 Lake Victoria and Lake Tanganyika responds more to DMI forcing, stretching as far as the tropical
510 rainforest in the Congo Basin (Figure 6).



511

512 *Figure 7. Spatial representations of the correlation between one-year accumulated SPI and indices*
513 *of regional climate phenomena highlight the areas where precipitation variability is sensitive to (a)*
514 *the El Niño Southern Oscillation and (b) the Indian Ocean Dipole Mode. Oceanic exposure and*
515 *local climate factors such as lakes partially predict sensitivity to these influences.*

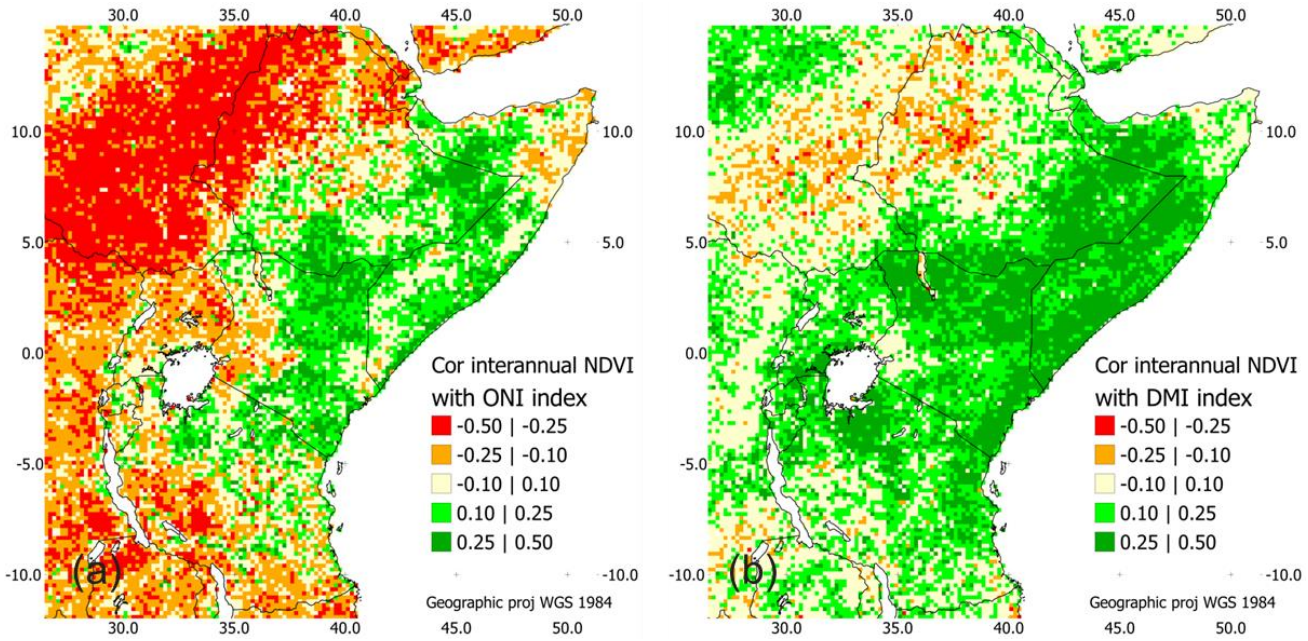



*Figure 8. Correlation maps of EEMD detected interannual NDVI with the historic ONI and DMI indices highlight the areas where vegetation changes are influenced by regional climate variability. These areas largely correspond to the areas where precipitation is forced by the same regional oceanic phenomena, demonstrating that meaningful interannual NDVI components can be extracted using EEMD decomposition.*

Next, it was assessed whether EEMD detected interannual NDVI is able to reveal the expected vegetation responses in the areas sensitive to climate variability. Figure 8 shows the strength of the coupling between interannual NDVI and the ONI and DMI indices respectively. The interannual NDVI follows the ONI index in most of the lowlands identified as sensitive to ENSO (Figure 8a), dominated by herbaceous and sparse shrub cover. However, the deciduous forest areas north of the Congo Basin do not respond to wet ENSO-driven episodes. As expected, vegetation in the continental lowlands responds negatively to lower than usual precipitation amounts during warm ENSO phases.

Correlation between interannual NDVI and DMI emphasize these findings. Although the Indian Ocean effects act differently throughout the study area, the detected responses of vegetation to DMI

episodes matches the areas identified as sensitive to DMI well in most areas. The ocean–exposed lowlands under low vegetation cover display greenness anomalies in response to increased precipitation during positive DMI episodes. The same holds for other DMI-sensitive areas in the Eastern Sahel and around Lake Victoria. Remarkably, the coupling of vegetation to the DMI episodes cannot be explained everywhere by its effect through precipitation, for example in vast areas of South Kenya and North Tanzania. Again, the forest ecosystems of the Congo Basin are neutral towards precipitation variability driven by Indian Ocean phenomena. A similar observation is made for the partly forested areas in South Tanzania: despite strongly DMI-forced precipitation, the NDVI response is rather ambiguous. Finally, the out of phase character of precipitation in the Ethiopian highlands with respect to DMI is also detected in the interannual NDVI.

The lower response of tropical forest ecosystems reflects the lower importance of water as a limiting factor, as well as the saturation effect of NDVI for high leaf area canopies (Sellers 1985). In addition, forest-climate interactions (Makarieva and Gorshkov 2007) and moisture recycling dynamics (van der Ent et al. 2010) are known to cause particular rainfall dynamics above forests, partly independent from the dominant oceanic moisture supply.

## 4. Discussion

Ensemble Empirical Mode Decomposition (EEMD) proves to be a suitable algorithm for the automated decomposition of large sets of NDVI time series into their characteristic time scales. The original EEMD decomposition technique (Huang et al. 1998; Wu and Huang 2009), complemented by the IMF sorting procedure proposed in §2.2 and clear choices of parameters for *sifting* and *splining* (Appendix A), is able to deliver the interannual components of any given time series.

With respect to reconstructing the interannual component from detected components, we reject the existing significance test to distinguish signal from noise-induced components for interannual

component detection, since neither the ‘a priori’ or ‘a posteriori’ version of the test yield better results than the ‘no test’ approach. Instead we weighted the capabilities of these significance tests to identify noise-induced low frequency components against the risk that these components may in fact contain signal as well as noise fragments (hereafter called *mixed* components). Although neither approach ensures a precise reconstruction of interannual components in terms of rRMSE, the linear and spectral properties of such components were best recovered by retaining all interannual IMFs (‘no test’). This implies that mixed components, if occurring in the decompositions of our simulated time series, contain on average more signal than noise fragments. Attempts to filter noise-affected components by subjecting individual IMFs to significance tests decreased performance measured by correlation and spectral coherence.

However, two nuances must be added to this finding. First, it may hold true for the tested ranges of $\sigma_{int}/\sigma_{ann}$ and absolute $\sigma_{noise}$, which were in turn based on estimated ranges from a real NDVI dataset. Outside these ranges noise may start to dominate low frequency EEMD components, prompting a significance test to retain the true signal components. This may be commonly the case for other biogeophysical or climatological datasets and applications, so our conclusion should be limited to the application on integrated time series of remotely sensed vegetation indices and its typical signal to noise ratios. Second, it can be argued that the purpose of IMF significance testing in a climatological context, as in the work of Coughlin and Tung (2005), is to isolate *pure* time scales with a high degree of probability to originate from signal rather than noise. This justifies strict filtering of potentially noise-induced IMFs. Conversely, when using multiple IMFs to reconstruct interannual components with an unknown combination of time scales, retaining both *pure* and *mixed* IMFs may be favorable as a strategy.

The values of the performance measures for given (inter)annual models and noise characteristics under the ‘no test’ approach (Table 5) give an account of the major sources of variability in performance, i.e., the relative strength of the interannual component compared to the dominant

annual cycle ($\sigma_{int}/\sigma_{ann}$), and the irregularity in timing of the annual season ($\varphi_{seas}$). It is worth noting that both factors do not act independently. The distorting effect of irregular growing seasons is of less impact when the superimposed interannual modulation is sufficiently high in amplitude, for example in arid ecoregions with a strong oceanic influence (Figure 6). A practical implication is that both factors must preferably be estimated upon detection. Low $\sigma_{int}/\sigma_{ann}$ ratios, corresponding to weak interannual changes compared to the NDVI range of the growing season, may hamper the detection and subsequent processing of a pixel's interannual component, unless the annual cycle displays a sufficiently regular timing. If threshold values for both parameters are considered, pixels can be masked out from further interpretation. Whereas the choice of thresholds is arbitrary and dependent on the further steps in a particular application, the results presented in Table 5 offer some guidelines. For example, a threshold of $[\sigma_{int}/\sigma_{ann} > 0.33]$ will generally ensure a good reproduction of the interannual component measured by correlation and spectral coherence. As seen from Table 1, this applies for all ecoregions with a dry or wet bimodal growing season. For unimodal (semi-arid to subhumid) ecoregions, the interannual NDVI components are buried under a more dominant seasonal component. However, if the estimated variability (measured as standard deviation) of the growing season peak position does not exceed half of the sample standard deviation used in our simulated prototype (0.10 years, see Table 2), then the performance level is largely maintained. We therefore recommend that automated use of EEMD to reveal interannual components from large gridded time series data (biogeophysical, climatological) be preceded by a calibration step, whereby preliminary estimates of the series' composition and characteristics are collected in order to assess the proper limits of detection.

A major implication for the further use of detected interannual components of NDVI (or other remotely sensed vegetation indices) in a causal analysis with gridded climate datasets, is the fact that detection performs relatively well when measured by correlation or spectral coherence, but proves imprecise with respect to its rRMSE deviation from the true component. This suggests that positive

606 and negative episodes in interannual vegetation change can be localized, and that the time scales
607 present in the series are represented well. Conversely, the high rRMSE values (~0.40 under optimal
608 conditions) indicate that an EEMD detected interannual component of a time series does not
609 necessarily represent the actual interannual NDVI change at each discrete point in time. Neither can
610 the absolute amplitudes of individual positive or negative episodes be interpreted as such. The power
611 of the decomposition tool lies in the capability to filter the desired time scales from a mixed signal,
612 and to produce a global representation of the occurrence and relative intensities of positive and
613 negative episodes. Further development of a technique to identify and isolate a climate-forced
614 component in the interannual NDVI component must take this knowledge into account. It must not
615 rely entirely on instantaneous responses in the time domain, but rather consider sufficiently large
616 time windows as well as the spectral domain for bivariate analysis with climate series.

617 A first effort to shift the EEMD analysis beyond the simulated setting towards real world NDVI
618 series demonstrated that spatially coherent patterns of vegetation dynamics can be revealed and
619 explained (partially) by episodes of regional climate oscillations. Gridded precipitation datasets -
620 remotely sensed as well as modeled using observational data - are expected to explain the detected
621 interannual vegetation dynamics at a more local scale. Eventually, explained interannual variability
622 can be separated from non-explained patterns, for further interpretative analysis over particular study
623 areas.

624 The case study on real NDVI and climate datasets over East and Central Africa demonstrates that
625 EEMD detection is able to produce meaningful indicators of interannual NDVI changes, as their
626 spatial distribution corresponds largely to what can be expected from empirical climate data. As for
627 any processing tool, the quality and the real world value of the output are inevitably limited by the
628 quality of the input data. In the case of the EEMD decomposition of NDVI time series,
629 inconsistencies in long term NDVI archives are the principal source of spurious patterns in the
630 detected interannual NDVI, which may jeopardize conclusions in thematic studies like the case study

631 in this paper. The most likely causes of inconsistency can be categorized as within-sensor
632 inconsistencies (e.g., orbital drift or imperfect corrections for it) or between-sensor inconsistencies
633 (e.g., historic processing differences, imperfect intercalibration of spectral response, or differences in
634 sun-target-sensor geometry). Other sources of error with a less systematic occurrence, such as sensor
635 noise and undetected clouds, will have less effect on the detected outcome, as the EEMD technique
636 efficiently isolates high frequent components. With this knowledge, the quantitative validation of the
637 detection tool was based on simulated data rather than on any existing real world dataset containing
638 residual inconsistencies. The LTDR/VGT dataset contains at least two potential sources of
639 inconsistency. The first one is the empirical linear sensor intercalibration between AVHRR and VGT
640 as described in Section 2.1.1 yields slightly biased residuals for NDVI values above 0.65. In practice,
641 this may cause an upward jump in NDVI at the 1999 breakpoint for pixels in densely forested areas.
642 Moreover, 1999 may not be representative for the full NDVI record in case of residual between-
643 sensor inconsistency in the LTDR version 2 data, which may cause an artifact interfering with the
644 former effect. The second source of inconsistency stems from between-sensor inconsistency of the
645 VGT-1 and VGT-2 sensors related to spectral response discrepancies. This effect has been identified
646 to cause an upward break in 2003 throughout all NDVI strata (Tian et al. 2015). For future
647 applications, alternative datasets can be considered. The LTDR version 4 provides improvement
648 upon the earlier version with respect to geolocation as well as an extension in time with successor
649 NOAA platforms (NASA 2014). The updated 15-day composite NDVI product of the Global
650 Inventory Modelling and Mapping Studies (GIMMS3g) results from efforts to recalibrate historic
651 NOAA-AVHRR data by correcting for effects introducing spurious NDVI trends such as orbital drift
652 and volcanic eruptions (Fensholt and Proud 2012; Pinzon et al. 2005). GIMMS3g is thus likely to
653 display improved stability over time, although potential artefacts have been identified over various
654 NDVI strata (Tian et al. 2015). Finally, the MODIS instruments have provided NDVI records from
655 1999 onwards, considered an improvement upon the AVHRR products due to their narrow band

656 definitions and wide dynamic range (Huete et al. 2002). Although generally deemed the alternative
657 dataset with highest consistency, the MODIS NDVI series was found to contain the effects of sensor
658 degradation (Wang et al. 2012).

659 Regardless the eventual choice of dataset, the EEMD decomposition and reconstruction approach
660 in its current form is deemed operational to conduct thematic follow-up research. It is
661 complementary to existing time series processing tools in the sense that 'interannual NDVI'
662 representing slow year-to-year evolution of vegetation greenness is a new key variable in remote
663 sensing-driven environmental research. Its user community may invoke extraction tools according to
664 their specific needs: BFAST (Verbesselt et al. 2010) for the detection of stepwise trends and
665 breakpoints relating to abrupt and gradual vegetation disturbances, such as wildfires and recovery;
666 TIMESAT (Jönsson and Eklundh 2002) for a detailed parameterization and quantification of
667 parameters related to the growing season; SPIRITS (Eerens et al. 2014) for near-real time
668 agricultural monitoring through processing of images and the extraction of vegetation indices related
669 to crop status. EEMD is proposed to complement this list of tools as one to extract interannual
670 vegetation changes in response to climatic oscillations and other sources of periodic disturbance
671 episodes.

672 Despite this applicability, some points for methodological improvement can be defined. A review
673 of existing methods to estimate a priori the nature and the level of noise from an NDVI time series
674 may contribute to the calibration phase necessary before EEMD application on a new dataset. If
675 noise can be modeled and estimated more accurately, the performance of the 'a posteriori'
676 significance test for IMFs may surpass that of the current 'no test' approach. Further refinement of
677 the EEMD application for automated analysis of gridded time series data can consist of evaluating
678 the effects of series length and temporal resolution, which were held constant at 1114 ten-day time
679 steps in this analysis.

680

## 5. Conclusions

Available archives of more than thirty years of remotely sensed optical imagery provide large spatiotemporal datasets, showing changes in condition of terrestrial vegetation as per-pixel NDVI trajectories. Beside seasonal variation, weaker but considerable interannual fluctuations exist in the NDVI series, and their identification is of high interest for research aiming at quantifying the long-term effects of climate variability on ecosystems. Thereto, signal processing techniques can be employed to detect interannual NDVI components within quantified limits of uncertainty.

We found that the Ensemble Empirical Mode Decomposition (EEMD) technique, extended with a sorting procedure for IMF components, is suitable to split a given NDVI time series into its constituent components with distinct time scales. The longer-than-annual EEMD components can be combined to reproduce the overall interannual oscillation present within the series. Implementation for automated processing of large NDVI datasets requires some complementary algorithmic formalization, which we proposed in this paper.

In order to gain insights in the performance of this method to accurately represent the true interannual component, and its sensitivities to data and model parameters, we tested its performance in a controlled setting with simulated NDVI time series. With respect to performance, we concluded that EEMD-detected interannual components generally represent well the occurrence of positive and negative episodes with proper time scales, whereas the local amplitudes must not be interpreted for individual episodes. This finding provides useful insights for subsequent bivariate analysis of interannual NDVI with precipitation series. With respect to sensitivities, the strength of the interannual component relative to the seasonal cycle, and modality and timing of the growing seasons, proved to be the principal factors affecting the detection performance. In addition, they interact and should therefore be assessed upon detection so that time series with low confidence in detection outcome can be excluded from the analysis.

The effectiveness of the EEMD decomposition method to detect interannual vegetation was demonstrated on real NDVI and climate data over East and Central Africa. In most areas where precipitation variability is sensitive to regional climatic and oceanic forcing, the method reveals interannual vegetation changes that show the expected responses, except in dense forest ecosystems. Finally, some methodological points of action were defined for further development of this tool to extract environmentally relevant information from growing remote sensing data sources.

**Acknowledgment**

This work was supported by the Flemish Institute for Technological Research (VITO) in Mol, Belgium through a PhD grant to Pieter Hawinkel, and by the Belgian Science Policy Office (Belspo) through the CVB contract with VITO. Stef Lhermitte was supported as postdoctoral researcher by Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO). Bruno Verbist was supported by the VLIR-UOS/ARES-CCD Acropolis platform KLIMOS on sustainable development funded by the Belgian Development Cooperation (DGD). We gratefully thank the anonymous reviewers for their thorough reading and constructive comments which helped to improve this article.

**References**

Achard, F., Eva, H.D., Stibig, H.-j., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J.-p. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science, 297*, 999-1002

Alston, W.N., Vaughan, S., & Uttley, P. (2013). The flux-dependent X-ray time lags in NGC 4051. *Monthly Notices of the Royal Astronomical Society, 435*, 1511-1519

Anyamba, A., & Eastman, J.R. (1996). Interannual variability of NDVI over Africa and its relation to El Nino Southern Oscillation. *International Journal of Remote Sensing, 17*, 2533-2548

727 Atzberger, C. (2013). Advances in remote sensing of agriculture: context description, existing
728 operational monitoring systems and major information needs. *Remote Sensing, 5*, 949-981

729 Bartholomé, E., & Belward, A.S. (2005). GLC2000: a new approach to global land cover mapping
730 from Earth observation data. *International Journal of Remote Sensing, 26*, 1959-1977

731 Baynard, C.W. (2013). Remote sensing applications: beyond land-use and land-cover change.
732 *Advances in Remote Sensing, 2*, 228-241

733 Brando, P.M., Goetz, S.J., Baccini, A., Nepstad, D.C., Beck, P.S.A., & Christman, M.C. (2010).
734 Seasonal and interannual variability of climate and vegetation indices across the Amazon.
735 *Proceedings of the National Academy of Sciences, 107*, 14685-14690

736 Brisson, E., Demuzere, M., Willems, P., & van Lipzig, N.M. (2015). Assessment of natural climate
737 variability using a weather generator. *Climate Dynamics, 44*, 495-508

738 Brown, M.E., de Beurs, K., & Vrieling, A. (2010). The response of African land surface phenology
739 to large scale climate oscillations. *Remote Sensing of Environment, 114*, 2286-2296

740 Camberlin, P., Martiny, N., Philippon, N., & Richard, Y. (2007). Determinants of the interannual
741 relationships between remote sensed photosynthetic activity and rainfall in tropical Africa. *Remote*
742 *Sensing of Environment, 106*, 199-216

743 Chen, C.-F., Son, N.-T., Chang, L.-Y., & Chen, C.-C. (2011). Monitoring of soil moisture variability
744 in relation to rice cropping systems in the Vietnamese Mekong Delta using MODIS data. *Applied*
745 *Geography, 31*, 463-475

746 Coughlin, K., & Tung, K.K. (2005). Empirical mode decomposition of climate variabilty in the
747 atmosphere. In N. Huang, & S. Shen (Eds.), *Hilbert-Huang Transform and Its Applications* (pp. 149-
748 166): World Scientific

749 Cracknell, A.P. (2001). The exciting and totally unanticipated success of the AVHRR in applications
750 for which it was never intended. *Advances in Space Research, 28*, 233-240

751 De Keersmaecker, W., Lhermitte, S., Tits, L., Honnay, O., Somers, B., & Coppin, P. (2015). A
752 model quantifying global vegetation resistance and resilience to short-term climate anomalies and
753 their relationship with vegetation cover. *Global Ecology and Biogeography, 24*, 539-548

754 Deronde, B., Debruyn, W., Gontier, E., Goor, E., Jacobs, T., Verbeiren, S., & Vereecken, J. (2014).
755 15 years of processing and dissemination of SPOT-VEGETATION products. *International Journal*
756 *of Remote Sensing, 35*, 2402-2420

757 Dierckx, W., Sterckx, S., Benhadj, I., Livens, S., Duhoux, G., Van Achteren, T., Francois, M.,
758 Mellab, K., & Saint, G. (2014). PROBA-V mission for global vegetation monitoring: standard
759 products and image quality. *International Journal of Remote Sensing, 35*, 2589-2614

760 Eerens, H., Haesen, D., Rembold, F., Urbano, F., Tote, C., & Bydekerke, L. (2014). Image time
761 series processing for agriculture monitoring. *Environmental Modelling & Software, 53*, 154-162

762 Feng, Z., Liang, M., Zhang, Y., & Hou, S. (2012). Fault diagnosis for wind turbine planetary
763 gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy
764 separation. *Renewable Energy, 47*, 112-126

765 Fensholt, R., Langanke, T., Rasmussen, K., Reenberg, A., Prince, S.D., Tucker, C., Scholes, R.J.,
766 Bao Le, Q., Bondeau, A., Eastman, R., Epstein, H., Gaughan, A.E., Hellden, U., Mbow, C., Olsson,
767 L., Paruelo, J., Schweitzer, C., Seaquist, J., & Wessels, K. (2012). Greenness in semi-arid areas

across the globe 1981–2007 — an Earth Observing satellite based analysis of trends and drivers. *Remote Sensing of Environment, 121*, 144-158

Fensholt, R., & Proud, S.R. (2012). Evaluation of Earth Observation based global long term vegetation trends - Comparing GIMMS and MODIS global NDVI time series. *Remote Sensing of Environment, 119*, 131-147

Flandrin, P., Rilling, G., & Goncalves, P. (2004). Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters, 11*, 112-114

Galford, G.L., Mustard, J.F., Melillo, J., Gendrin, A., Cerri, C.C., & Cerri, C.E.P. (2008). Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sensing of Environment, 112*, 576-587

Guan, B.T. (2014). Ensemble empirical mode decomposition for analyzing phenological responses to warming. *Agricultural and Forest Meteorology, 194*, 1-7

Guo, L., Wu, S., Zhao, D., Yin, Y., Leng, G., & Zhang, Q. (2014). NDVI-based vegetation change in Inner Mongolia from 1982 to 2006 and its relationship to climate at the biome scale. *Advances in Meteorology, 2014*, 12

Guttman, N.B. (1999). Accepting the Standardized Precipitation Index: a calculation algorithm. *Journal of the American Water Resources Association, 35*, 311-322

Hilker, T., Lyapustin, A.I., Tucker, C.J., Hall, F.G., Myneni, R.B., Wang, Y.J., Bi, J., de Moura, Y.M., & Sellers, P.J. (2014). Vegetation dynamics and rainfall sensitivity of the Amazon. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 16041-16046

788 Hirota, M., Holmgren, M., Van Nes, E.H., & Scheffer, M. (2011). Global Resilience of Tropical
789 Forest and Savanna to Critical Transitions. *Science, 334*, 232-235

790 Holmgren, M., Hirota, M., van Nes, E.H., & Scheffer, M. (2013). Effects of interannual climate
791 variability on tropical tree cover. *Nature Climate Change, 3*, 755-758

792 Holmgren, M., Scheffer, M., Ezcurra, E., Gutiérrez, J., & Mohren, G.F.J. (2001). El Niño effects on
793 the dynamics of terrestrial ecosystems. *Trends in Ecology & Evolution, 16*, 89-94

794 Hsu, K.-l., Gupta, H.V., Gao, X., & Sorooshian, S. (1999). Estimation of physical variables from
795 multichannel remotely sensed imagery using a neural network: Application to rainfall estimation.
796 *Water Resources Research, 35*, 1605-1618

797 Huang, B., Banzon, V.F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T.C., Smith, T.M., Thorne,
798 P.W., Woodruff, S.D., & Zhang, H.-M. (2014). Extended Reconstructed Sea Surface Temperature
799 Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. *Journal of Climate, 28*, 911-930

800 Huang, N.E., Shen, Z., & Long, S.R. (1999). A new view on nonlinear water waves. *Annual Review*
801 *of Fluid Mechanics, 31*, 417-457

802 Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., & Liu,
803 H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-
804 stationary time series analysis. *Proceedings of the Royal Society of London, 454*, 903-995

805 Huang, N.E., Wu, M.C., Long, S.R., Shen, S.S., Wendong, Q., Gloersen, P., & Fan, K.L. (2003). A
806 confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of*
807 *the Royal Society of London, 459*, 2317-2345

808 Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., & Ferreira, L.G. (2002). Overview of the
809 radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of*
810 *Environment, 83,* 195-213

811 Immerzeel, W.W., Quiroz, R.A., & De Jong, S.M. (2005). Understanding precipitation patterns and
812 land use interaction in Tibet using harmonic analysis of SPOT VGT-S10 NDVI time series.
813 *International Journal of Remote Sensing, 26,* 2281-2296

814 IPCC (2007). Climate Change 2007: The Physical Science Basis. Contribution of Working Group I
815 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In S. Solomon,
816 D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, & H.L. Miller (Eds.) (p. 996):
817 Intergovernmental Panel on Climate Change

818 JAMSTEC (2010). Indian Ocean Dipole Mode Index (DMI). Japan Agency for Marine-Earth
819 Science and Technology. In J.A.f.M.-E.S.a. Technology (Ed.)

820 Ji, L., & Gallo, K. (2006). An agreement coefficient for image comparison. *Photogrammetric*
821 *Engineering & Remote Sensing, 72,* 823-833

822 Jönsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of
823 satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing, 40,* 1824-1832

824 Kuo, C.-Y., Wei, S.-K., & Tsai, P.-W. (2013). Ensemble Emprirical Mode Decomposition with
825 supervised cluster analysis. *Advances in Adaptive Data Analysis, 5*

826 Lhermitte, S., Verbesselt, J., Jonckheere, I., Nackaerts, K., van Aardt, J.A.N., Verstraeten, W.W., &
827 Coppin, P. (2008). Hierarchical image segmentation based on similarity of NDVI time series. *Remote*
828 *Sensing of Environment, 112,* 506-521

829 Lhermitte, S., Verbesselt, J., Verstraeten, W.W., & Coppin, P. (2011). A comparison of time series
830 similarity measures for classification and change detection of ecosystem dynamics. *Remote Sensing*
831 *of Environment, 115*, 3129-3152

832 Liu, M., Tian, H., Chen, G., Ren, W., Zhang, C., & Liu, J. (2008). Effects of Land-Use and Land-
833 Cover Change on Evapotranspiration and Water Yield in China During 1900-2001. *JAWRA Journal*
834 *of the American Water Resources Association, 44*, 1193-1207

835 Luo, Y., Melillo, J., Niu, S., Beier, C., Clark, J.S., Classen, A.T., Davidson, E., Dukes, J.S., Evans,
836 R.D., Field, C.B., Czimczik, C.I., Keller, M., Kimball, B.A., Kueppers, L.M., Norby, R.J., Pelini,
837 S.L., Pendall, E., Rastetter, E., Six, J., Smith, M., Tjoelker, M.G., & Torn, M.S. (2011). Coordinated
838 approaches to quantify long-term ecosystem dynamics in response to global change. *Global Change*
839 *Biology, 17*, 843-854

840 Maisongrande, P., Duchemin, B., & Dedieu, G. (2004). VEGETATION/SPOT: an operational
841 mission for the Earth monitoring; presentation of new standard products. *International Journal of*
842 *Remote Sensing, 25*, 9-14

843 Makarieva, A.M., & Gorshkov, V.G. (2007). Biotic pump of atmospheric moisture as driver of the
844 hydrological cycle on land. *Hydrology and Earth System Sciences, 11*, 1013-1033

845 Mann, M.E., & Lees, J.M. (1996). Robust estimation of background noise and signal detection in
846 climatic time series. *Climatic Change, 33*, 409-445

847 Martínez, B., & Gilabert, M.A. (2009). Vegetation dynamics from NDVI time series analysis using
848 the wavelet transform. *Remote Sensing of Environment, 113*, 1823-1842

849 Molla, M.K.I., Ghosh, P.R., & Hirose, K. (2011). Bivariate EMD-based data adaptive approach to
850 the analysis of climate variability. *Discrete Dynamics in Nature and Society, 2011*, 21

851 Myneni, R.B., Los, S.O., & Tucker, C.J. (1996). Satellite-based identification of linked vegetation
852 index and sea surface temperature anomaly areas from 1982-1990 for Africa, Australia and South
853 America. *Geophysical research letters, 23*, 729-732

854 Nagol, J.R., Vermote, E.F., & Prince, S.D. (2014). Quantification of impact of orbital drift on inter-
855 annual trends in AVHRR NDVI data. *Remote Sensing, 6*, 6680-6687

856 NASA (2014). Land Long Term Data Record. In, *URL http://ltdr.nascom.nasa.gov/cgi-*
857 *bin/ltdr/ltdrPage.cgi?fileName=LTDR_update, last accessed 24 June 2015.*

858 Nicholson, S.E., & Farrar, T.J. (1994). The influence of soil type on the relationships between NDVI,
859 rainfall, and soil moisture in semiarid Botswana. I. NDVI response to rainfall. *Remote Sensing of*
860 *Environment, 50*, 107-120

861 Nicholson, S.E., & Kim, J. (1997). The relationship of the El Niño-Southern Oscillation to African
862 rainfall. *International Journal of Climatology, 17*, 117-135

863 NOAA-CPC (2014). Historic El Niño/La Niña episodes (1950-present). In, *URL*
864 *http://www.cpc.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml, last accessed 24*
865 *June 2015. Climate Prediction Center, National Oceanic and Atmospheric Administration*. Climate
866 Prediction Center, National Oceanic and Atmospheric Administration

867 Pedelty, J., Devadiga, S., Masuoka, E., Brown, M., Pinzon, J., Tucker, C., Roy, D., Ju, J., Vermote,
868 E., Prince, S., Nagol, J., Justice, C., Schaaf, C., Liu, J., Privette, J., & Pinheiro, A. (2007). Generating
869 a long-term land data record from the AVHRR and MODIS instruments. In, *IEEE International*
870 *Geoscience and Remote Sensing Symposium (IGARSS), 23-28 July 2007* (pp. 1021-1025). Barcelona,
871 Spain: IEEE

872 Pegram, G.G.S., Peel, M.C., & McMahon, T.A. (2008). Empirical mode decomposition using
873 rational splines: an application to rainfall time series. *Proceedings of the Royal Society A, 464*, 1483–
874 1501

875 Piao, S., Nan, H., Huntingford, C., Ciais, P., Friedlingstein, P., Sitch, S., Peng, S., Ahlström, A.,
876 Canadell, J.G., Cong, N., Levis, S., Levy, P.E., Liu, L., Lomas, M.R., Mao, J., Myneni, R.B., Peylin,
877 P., Poulter, B., Shi, X., Yin, G., Viovy, N., Wang, T., Wang, X., Zaehle, S., Zeng, N., Zeng, Z., &
878 Chen, A. (2014). Evidence for a weakening relationship between interannual temperature variability
879 and northern vegetation activity. *Nature Communications, 5*

880 Piao, S.L., Wang, X.H., Ciais, P., Zhu, B., Wang, T., & Liu, J. (2011). Changes in satellite-derived
881 vegetation growth trend in temperate and boreal Eurasia from 1982 to 2006. *Global Change Biology,*
882 *17*, 3228-3239

883 Pinzon, J.E., Brown, M.E., & Tucker, C.J. (2005). EMD correction of orbital drift artifactsin satellite
884 data stream. In N.E. Huang, & S.S.P. Shen (Eds.), *Hilbert-Huang transform and its applications* (pp.
885 167-186): World Scientific Publishing

886 Plisnier, P.D., Serneels, S., & Lambin, E.F. (2000). Impact of ENSO on East African ecosystems: a
887 multivariate analysis based on climate and remote sensing data. *Global Ecology and Biogeography,*
888 *9*, 481-497

889 Quiroz, R., Yarlequé, C., Posadas, A., Mares, V., & Immerzeel, W.W. (2011). Improving daily
890 rainfall estimation from NDVI using a wavelet transform. *Environmental Modelling & Software, 26*,
891 201-209

892 Richard, Y., & Poccard, I. (1998). A statistical study of NDVI sensitivity to seasonal and interannual
893 rainfall variations in Southern Africa. *International Journal of Remote Sensing, 19*, 2907-2920

894 Roerink, G.J., Menenti, M., Soepboer, W., & Su, Z. (2003). Assessment of climate impact on
895 vegetation dynamics by using remote sensing. *Physics and Chemistry of the Earth, 28*, 103–109

896 Sellers, P.J. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of*
897 *Remote Sensing, 6*, 1335-1372

898 Sitters, J., Holmgren, M., Stoorvogel, J.J., & López, B.C. (2012). Rainfall-tuned management
899 facilitates dry forest recovery. *Restoration Ecology, 20*, 33-42

900 Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H.V., Imam, B., & Braithwaite, D. (2000). Evaluation of
901 PERSIANN System Satellite–Based Estimates of Tropical Rainfall. *Bulletin of the American*
902 *Meteorological Society, 81*, 2035-2046

903 Steven, M.D., Malthus, T.J., Baret, F., Xu, H., & Chopping, M.J. (2003). Intercalibration of
904 vegetation indices from different sensor systems. *Remote Sensing of Environment, 88*, 412-422

905 Swets, D.L., Reed, B.C., Rowland, J.D., & Marko, S.E. (1999). A weighted least-squares approach to
906 temporal NDVI smoothing. In, *ASPRS Annual Conference, 17-21 May 1999* (pp. 526-536). Oregon,
907 Portland: American Society for Photogrammetry and Remote Sensing

908 Swinnen, E. (2008). Vegetation dynamics in Southern Africa from NOAA-AVHRR and SPOT-VGT
909 time series. In, *Faculté des sciences. École doctorale geosciences.* (p. 208). Louvain-la-Neuve:
910 Université catholique de Louvain

911 Swinnen, E., & Veroustraete, F. (2008). Extending the SPOT-VEGETATION NDVI Time Series
912 (1998–2006) Back in Time With NOAA-AVHRR Data (1985–1998) for Southern Africa. *IEEE*
913 *Transactions on Geoscience and Remote Sensing, 46*, 558-572

914 Tarnavsky, E., Garrigues, S., & Brown, M.E. (2008). Multiscale geostatistical analysis of AVHRR,
915 SPOT-VGT, and MODIS global NDVI products. *Remote Sensing of Environment, 112*, 535-549

916 Tian, F., Fensholt, R., Verbesselt, J., Grogan, K., Horion, S., & Wang, Y. (2015). Evaluating
917 temporal consistency of long-term global NDVI datasets for trend analysis. *Remote Sensing of*
918 *Environment, 163*, 326-340

919 Torrence, C., & Compo, G.P. (1998). A practical guide to wavelet analysis. *Bulletin of the American*
920 *Meteorological Society, 79*, 61-78

921 Torrence, C., & Webster, P.J. (1999). Interdecadal Changes in the ENSO–Monsoon System. *Journal*
922 *of Climate, 12*, 2679-2690

923 Trishchenko, A.P., Cihlar, J., & Li, Z. (2002). Effects of spectral response function on surface
924 reflectance and NDVI measured with moderate resolution satellite sensors. *Remote Sensing of*
925 *Environment, 81*, 1-18

926 Tucker, C.J., & Sellers, P.J. (1986). Satellite remote sensing of primary production. *International*
927 *Journal of Remote Sensing, 7*, 1395-1416

928 van der Ent, R.J., Savenije, H.H.G., Schaefli, B., & Steele-Dunne, S.C. (2010). Origin and fate of
929 atmospheric moisture over continents. *Water Resources Research, 46*, n/a-n/a

930 Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010). Phenological change detection
931 while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of*
932 *Environment, 114*, 2970-2980

933 Vermote, E., Justice, C.O., & Breon, F.M. (2009). Towards a Generalized Approach for Correction
934 of the BRDF Effect in MODIS Directional Reflectances. *Geoscience and Remote Sensing, IEEE*
935 *Transactions on, 47*, 898-908

936 Vermote, E., & Kaufman, Y.J. (1995). Absolute calibration of AVHRR visible and near-infrared
937 channels using ocean and cloud views. *International Journal of Remote Sensing, 16*, 2317-2340

938 Wang, D., Morton, D., Masek, J., Wu, A., Nagol, J., Xiong, X., Levy, R., Vermote, E., & Wolfe, R.
939 (2012). Impact of sensor degradation on the MODIS NDVI time series. *Remote Sensing of*
940 *Environment, 119*, 55-61

941 Williams, C.A., & Hanan, N.P. (2011). ENSO and IOD teleconnections for African ecosystems:
942 evidence of destructive interference between climate oscillations. *Biogeosciences, 8*, 27-40

943 Wu, Z., & Huang, N.E. (2004). A study of the characteristics of white noise using the empirical
944 mode decomposition method. *Proceedings of the Royal Society of London, 460*, 1597-1611

945 Wu, Z., & Huang, N.E. (2005). Statistical significance test of Intrinsic Mode Functions. In N.E.
946 Huang, & S.S.P. Shen (Eds.), *Hilbert-Huang transform and its applications* (pp. 107-127): World
947 Scientific

948 Wu, Z., & Huang, N.E. (2009). Ensemble Empirical Mode Decomposition: A noise-assisted data
949 analysis method. *Advances in Adaptive Data Analysis, 1*, 1-41

950 Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review.
951 *Journal of Plant Ecology, 1*, 9-23

952 Zhou, H., Van Rompaey, A., & Wang, J.a. (2009). Detecting the impact of the “Grain for Green”
953 program on the mean annual vegetation cover in the Shaanxi province, China using SPOT-VGT
954 NDVI data. *Land Use Policy, 26*, 954-960

955 Zhou, L., Tian, Y., Myneni, R.B., Ciais, P., Saatchi, S., Liu, Y.Y., Piao, S., Chen, H., Vermote, E.F.,
956 Song, C., & Hwang, T. (2014a). Widespread decline of Congo rainforest greenness in the past
957 decade. *Nature, 509*, 86-90

958 Zhou, Q., Jiang, H., Wang, J., & Zhou, J. (2014b). A hybrid model for PM2.5 forecasting based on
959 ensemble empirical mode decomposition and a general regression neural network. *Science of The*
960 *Total Environment, 496*, 264-274
961
962

963 **Appendix A. The Empirical Mode Decomposition algorithm (Huang et al. 1998)**

964 A series X(t) is decomposed into a finite number of Intrinsic Mode Functions (IMFs) and a
965 residual term R(t) following the additive model (Eq.3).

966 The locations and values of all local minima and maxima in a series X(t) are detected using an
967 extrema-finding function. Endpoints of the series are not considered as extrema unless they lie
968 outside the range defined by the nearest minimum and maximum. For flat sections, the (left) middle
969 location is returned.

970 The lower (upper) envelope of X(t) is obtained by connecting the local minima (maxima) of X(t)
971 using a cubic spline. Cubic splines are uniquely defined when values for the second derivatives at the
972 endpoints are provided. We follow the two-step approach of Pegram *et al*. (2008), here described for
973 the left end of the lower (upper) envelope. The first original minimum (maximum) is mirrored
974 around the left endpoint, and assigned the same second derivative. A cubic spline is calculated

975 through the minima (maxima), ignoring the left endpoint. If the left endpoint is found to lie outside
976 the calculated envelopes, it is added to the extrema set and the spline is recalculated. This is to avoid
977 splines to vary freely at the ends, which would cause end effects propagating through the following
978 iterative sifting procedure.

979 The mean of lower and upper envelopes M(t) contains the low-frequent information of X(t). The
980 first ProtoMode Function ($PMF_1$) of X(t) is obtained by subtracting M(t) from X(t). This step is
981 termed 'sifting' and $PMF_1$ is enriched in high frequency information. Sifting $PMF_1$ yields $PMF_2$ and
982 the procedure is repeated *k* times until the number of extrema and the number of zero-crossings have
983 not changed for *S* consecutive siftings. $PMF_k$ is taken as the first Intrinsic Mode Function ($IMF_1$) of
984 X(t). We follow the guidelines by Huang *et al*. (2003) and use [S = 5] as a stopping criterion for the
985 sifting process.

986 The difference of X(t) and $IMF_1$ is the (preliminary) residual. It is ingested in the sifting
987 procedure to yield $IMF_2$, which will contain the high frequency information after subtraction of
988 $IMF_1$. More IMFs are obtained iteratively with increasing frequencies, until the (final) residual R(t)
989 is monotonic or has one extremum. It can be regarded as the residual trend in the series.

990 The computing cost of the EMD algorithm increases quadratically with the length $N_t$ of the time
991 series. Its most costly step is the cubic spline interpolation, which involves manipulations on $N_t \; x \; N_t$
992 matrices. For $N_t = 1114$, the spline interpolation accounts for nearly 80% of the total computing
993 cost of 1163 msec. In our IDL implementation, the computing time $t_{comp}$ (in msec) for one series of
994 length $N_t$ is on average given by

995 $$t_{comp} = (0.000858 * {N_t}^2 + 0.282 * N_t + 23.2) \qquad \text{(Eq.A1)}$$

996 For the EEMD extension, there is an additional linear increase with the number of noise-added
997 iterations (set to 50 in this study). The vast number of pixels in a spatio-temporal dataset as the image

998 sets over East and Central Africa (>14000 land pixels at 20 km resolution) further pushes the total
999 processing time into the order of magnitude of weeks.

1000 **Appendix B. Significance test for IMFs based on period-energy relationships of white noise**

1001 The theoretical and derived empirical relationships to distinguish signal from noise, are based on
1002 two key characteristics of an IMF of a series with length $N_t$: its energy density

1003 $$E = \frac{1}{N_t}\sum IMF^2 \qquad \text{(Eq.B1)}$$

1004 and its theoretical mean period *T*, estimated as $\widehat{\mathrm{T}}$ by counting the number of extrema divided by
1005 twice the series length. Wu and Huang derived that the energy of a population of white noise series
1006 normalized to unit energy density has a mean $\overline{\mathrm{E}}$, related to its mean period *T*:

1007 $$\ln(\bar{E}) = -\ln(T) \qquad \text{(Eq.B2)}$$

1008 Moreover, the probability distribution of the energy density of a white noise IMF can be derived
1009 from the finding that $N_t * \mathrm{E}$ follows a $\chi^2$-distribution with $N_t * E$ degrees of freedom:

1010 $$p(N_t * E) = (N_t * E)^{N_t * \bar{E}/2-1} \exp\left(-\frac{N_t * E}{2}\right) / \left(2^{N_t * \bar{E}/2}\Gamma(N_t * \bar{E}/2)\right) \qquad \text{(Eq.B3)}$$

1011 Through log-transformation of the energy density

1012 $$y = \ln(E) \qquad \text{(Eq.B4)}$$

1013 and substitution of Equations B2 and B4 into the probability distribution of Equation B3, the
1014 confidence levels for the log-energy density of an IMF originating from white noise can be estimated
1015 from its mean period $\hat{T}$.

1016 $$p(y) = (N_t * \exp(y))^{N_t/(2\hat{T})} \exp(-N_t * \exp(y)/2) \,/\, \left(2^{N_t/(2\hat{T})}\Gamma\left(N_t/(2\hat{T})\right)\right) \qquad \text{(Eq. B5)}$$

1017 The bias in the estimation of $\hat{T}$ affects Equation B2 but preserves its linear relationship. Monte
1018 Carlo verification by a large number of random white noise simulations yielded the empirical
1019 relationship for series length [$N_t$ = 1114]:

1020 $$\ln(\bar{E}) = -1.08\ln(T) + 0.12 \qquad \text{(Eq.B6)}$$

1021 The energy distributions and confidence levels were recalculated accordingly, and verified with
1022 the simulation result. The (shifted) theoretical 99% confidence limit proved to bound 95% of the
1023 white noise simulations. Similar outcomes with slightly different coefficients were found for red
1024 noise simulations with [AC1 = {0.25, 0.50}].
1025 As white noise is assumed to contain no information, real signal components will display
1026 energies higher than can be expected from white noise with a chosen level of confidence. Under the
1027 'a priori' approach, a given time series consisting of signal and noise is normalized to unit energy
1028 density and treated as being potentially pure noise. If however the noise level in the series can be
1029 estimated, the normalization step can be performed more accurately 'a posteriori' and more signal
1030 components can be distinguished. Both approaches were evaluated in this paper.
1031

**List of Figure Captions**

Figure 1. The Oceanic Niño Index (ONI) and the Dipole Mode Index (DMI) are derived from sea temperature measurements and represent the positive and negative episodes of climate oscillations that affect interannual rainfall patterns over East and Central Africa (IPCC 2007).

Figure 2. A temporal NDVI profile (pixel 7 in Figure 3, North Tanzania) and its Intrinsic Mode Functions by (a) EMD decomposition and (b) the improved EEMD decomposition. EEMD decreases 'mode mixing' whereby fragments of a single time scale occur in another component, e.g. for the annual time scale for the period 1984-1986 (box on graph (a)).

Figure 3. Mean NDVI value over the period 1981-2012 per image pixel over East and Central Africa. The square symbols indicate the locations of the selected series pixels(Figure 1), used to estimate parameters for NDVI time series simulation (Table 1).

Figure 4. Time series of the Normalized Difference Vegetation Index (NDVI) for 9 pixels in distinct climate zones in East and Central Africa (Figure 3), derived from cross-calibrated image data from the NOAA-AVHRR (1981-1999) and SPOT-VGT (1998-2012) satellite sensors. Three prototypes of growing seasons were selected for data simulations: (*)unimodal (one rainy season) and (**)bimodal (two rainy seasons) with either overall low (L) or high (H) NDVI levels.

Figure 5. Overview of the evaluation framework. Mixtures X(t) of simulated annual (ANN), interannual (INT) and noise components are decomposed using EEMD. The interannual component

(red solid line) is reconstructed from selected components and evaluated against the input component (blue dashed line).

Figure 6. The distribution of land cover types over the study area (GLC2000; Bartholomé and Belward 2005) reflects to a large extent the different levels of mean NDVI shown in Figure 3. Ecosystems respond differently to regional precipitation variability, as confirmed in the demonstration case study over East and Central Africa (see Discussion section).

Figure 7. Spatial representations of the correlation between one-year accumulated SPI and indices of regional climate phenomena highlight the areas where precipitation variability is sensitive to (a) the El Niño Southern Oscillation and (b) the Indian Ocean Dipole Mode. Oceanic exposure and local climate factors such as lakes partially predict sensitivity to these influences.

Figure 8. Correlation maps of EEMD detected interannual NDVI with the historic ONI and DMI indices highlight the areas where vegetation changes are influenced by regional climate variability. These areas largely correspond to the areas where precipitation is forced by the same regional oceanic phenomena, demonstrating that meaningful interannual NDVI components can be extracted using EEMD decomposition.